REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author DILFONI, A.

COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOAN

Theses may not be lent to individuals, but the University Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: The Theses Section, University of London Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the University of London Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The University Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

☑ This copy has been deposited in the Library of UCL.

☐ This copy has been deposited in the University of London Library, Senate House, Malet Street, London WC1E 7HU.

DNA binding specificities and transcriptional activity of Mirror, an Iroquois transcription factor in *Drosophila melanogaster.*

Aphrodite Bilioni

Developmental Patterning Laboratory

Cancer Research UK, London

A thesis submitted for the degree of

Doctor of Philosophy

at the University of London

September 2004

UMI Number: U591851

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U591851
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Iroquois family genes regulate diverse developmental processes in higher eukaryotes. The *Drosophila Iroquois* complex consists of three closely related genes: *mirror (mirr), araucan (ara) and caupolican (caup). Iroquois* genes have been shown to be essential in various systems for processes such as neural development, patterning of the vertebrate heart, establishment of planar cell polarity and axon pathfinding. However, despite extensive work demonstrating the importance of Iroquois transcription factors in diverse aspects of development, it is still unclear how they act to control transcription or, which are their downstream targets.

Iroquois proteins are atypical Homeodomain (HD) transcription factors of the TALE class. To determine the DNA-binding preference of the Iroquois family members I conducted a DNA binding site selection experiment using *Drosophila* Mirror and defined a novel consensus sequence ACAnnTGT, which is different from the classic HOX motif. I provide evidence that this motif is the minimum requirement for Mirror binding to DNA. Other members of the Iroquois family can recognise this site *in vitro* suggesting that it may be a universal Iroquois binding site (IBS). Mirror binds the IBS as a homodimer and can also form heterodimers with other Iroquois proteins.

To test if the site is functional I made *in vivo* reporter constructs and showed that the site can mediate transcriptional repression in transgenic flies. Genome-wide searches for occurrences of transcription factor binding sites can help identifying their downstream targets. In combination with a microarray screen, recently carried out in our lab to identify Mirror downstream targets, we have looked for Mirror binding sites in genomic regions of candidate genes. The case of a candidate for direct Mirror target will be presented.

Acknowledgements

This is it... It took four years and a lot of work, has given me new skills, new friends, has made London a place I'll always carry in my heart and has finally reached an end. There are various people who share a great part of the credits and this is where I get to acknowledge their help: Without you this would not have been possible.

First of all I would like to thank my supervisor Helen McNeill. She has been a constant source of ideas, encouragement and advice and has always been there when I needed to talk. Her genuine enthusiasm about science has always been a great motivation for me. I should also thank my second supervisor Caroline Hill and Mike Howell in her lab, for teaching me everything I know about transcription and "bandshifts" and for the long conversations about exciting, surprising and sometimes incomprehensible results.

Then there are, of course, the people in the lab. Starting with the old crew, also known as "the collective". The "accomplices" in my first years of "mirroring", Ina and Mike, great working partners and great friends. Hope the Mirror bond will never break!!! Manolito, thanks for all the (fly-) wisdom and the versatile transferable skills (microwaves will never look the same!). Won't miss the music though!!! I will always remember Kirsten with a lot of love, I feel lucky to have known her, I just wish I had told her too... To the current "McNeills", Liz, Afifa, Nadja, Joe, Ruth, Sakura and Lesley. It's been a great pleasure working with all of you. I am grateful for all your help and for making every day in the lab such a pleasant experience. A great big "thank you" to Terence and Steve, for the injections and for taking such good care of the flies! To all the members of the developmental labs at CR-UK, for sharing ideas, reagents, for their help and advice and for the innumerable pints of beer (this last one goes mainly for "little" Julian, Alex, Simon and Paquito).

To my "family" in London, Brunakis, Nickolakis, Joaninha, Oscar, Sofia, Highlander and Paquito for the unforgettable "Golders Green days". My fondest memories of all these years are marked by your smiles...thank you. To Σοφουλα and Don for being as close to parents as one could ask. And to my climbing partners for showing me the views from the top and the many hours of fun (and sweat...) we've spent together.

Moving south, where it all began. To my family, my parents and my brother, for their unlimited love and support, for the greatest lessons in life and for sharing every part of the way. I can never thank you enough.

To all my friends in Greece for staying "so close no matter how far", for all the phone calls, messages, letters and visits and for making me feel like I never left home. Ελενη, Χρηστο, Στελιο, Μαρια, Ελενη, Αργυρη, Τασο,Κλεοπατρα, Σταυρουλα, Βασιλη, Αγγελικουλα , Φανη, Γιαννη και Γεωργια.....το the past, the present and the future!

Finally to Alejandro, for the words, the music and the silences; for eliminating the distance, sharing the dream and embarking on this beautiful journey. Gracias, mi sol...

Στην οικογενεια μου και στους φιλους που θα μεινουν για παντα

To my family and to friendships that last forever

A mi familia y a los amigos que estarán ahí siempre

Table of Contents

Table of Figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

List of abbreviations

This is a list of the abbreviations used throughout the document. In all cases the full name and the appropriate abbreviation are stated the first time each one of these terms is used within the text

β-gal	Beta galactosidase
antp	*antennapedia*
ac-sc	*achaete-scute*
ara	*araucan*
b-HLH	Basic Helix-Loop-Helix
BR	Basic Region
bx	*bithorax*
CNS	Central Nervous System
caup	*caupolican*
comm	*commissureless*
DE	Distal Enhancer (in the *goosecoid* promoter)
Dfd	*deformed*
DV axis	Dorsal Ventral axis
EMSA	Electrophoretic Mobility Shift Assay
exd	*extradenticle*
Gbe	Grainyhead Binding Element
HD	Homeodomain
HOX	Homeotic gene Complex
hth	*homothorax*
IBS	Iroquois Binding Site
IC	Information Content
IP	Immunoprecipitation
Iro, Irx	*Iroquois*
kr	*krüppel*
MEME	Multiple Expectation/Maximisation algorithm for Motif Elicitation
mirr	*mirror*
N	Notch
NB	neuroblast
o/n	overnight
ORF	Open Reading Frame
Pc-G	Polycomb Group
PCR	Polymerase Chain Reaction
PWM	Position Weight Matrix
rho	*rhomboid*
rt	room temperature
Scr	*Sex combs reduced*
shh	Sonic hedgehog
SMC	Sensory Mother Cell
TALE	Three Amino acid Loop Extension
Xgal	5-bromo-4-cloro-3-indolyl-β-galactopyranoside
Xiro	Xenopus Iroquois
Ziro	Zebrafish Iroquois

Chapter 1: Introduction

1.1. Homeotic genes: origin and evolution

Homeotic genes are master control genes that specify body parts and regulate developmental processes. They were first studied in *Drosophila* and their name refers to the fact that loss of function mutations of these genes cause transformations of one body structure into another (homeo is the Greek word for identical). The first homeotic mutant reported in the literature was *bithorax,* discovered by Bridges in 1915 as a transformation of the *Drosophila* metathorax (T3) towards mesothorax (T2) (reviewed in Duncan and Montgomery, 2002a; Duncan and Montgomery, 2002b). In 1966 W. Gehring reported a homeotic mutation that transformed the antennae on the head of the fly into a pair of middle legs. This mutation, originally named Nasobemia turned out to be a dominant gain-of-function mutation at what is now known as the *Antennapedia* locus. Homeotic genes have since been shown to form a conserved genetic network for patterning the anteroposterior axis in all bilateral animals (McGinnis and Krumlauf, 1992).

Homeotic genes are organized in clusters and it is believed that they arose by tandem duplication of an ancestral gene. In *Tribolium* (red flour beetle) there is a single homeotic gene complex (HOX) (Beeman, 1987). In *Drosophila* the cluster has split in two generating the *Bithorax* (*BX-C*) and the *Antennapedia* (*ANTP-C*) complex (Lewis, 1978). In mice and humans there are four clusters (*HoxA* to *HoxD*) located on

four different chromosomes. Early in the evolution of vertebrates (at least 500million years ago) a duplication occurred that gave rise to two clusters. Later in vertebrate evolution the two clusters duplicated again, most probably through an entire genome duplication to form the four clusters A, B, C and D (reviewed in Prince and Pickett, 2002) (Figure 1.1). The corresponding genes in each cluster are therefore considered to be paralogues. In zebrafish there are as many as 7 *Hox* clusters suggesting that an additional genome duplication took place in the lineage that led to teleost fish (Amores *et al.*, 1998). At the opposite end of the evolutionary ladder *C.elegans* has a reduced, dispersed and less conserved *Hox* gene complement indicating that nematodes are on a separate branch of the evolutionary tree (reviewed in Aboobaker and Blaxter, 2003).

Within the cluster it has been reported that the most centrally placed genes (i.e *Antp* in *Drosophila* and its counterparts in the vertebrate *Hox* gene clusters, namely *Hox6* and *7)* deviate the least from the consensus sequence of the Homeodomain. The degree of divergence increases as we move to genes located more 5' and 3' in the cluster so that the genes *labial (Hox 1)* and *Abdominal B (Hox13)* at the two termini of the cluster contain the most divergent Homeodomains (Gehring *et al.*, 1994a). This observation suggests that the primordial clusters may have arisen through a series of duplications, the first of which resulted in the generation of the two terminal genes. The more internal genes are later additions in the cluster and have therefore had less time to diverge.

As first observed in *Drosophila*, homeotic genes have another interesting feature. Their position within the cluster corresponds to their time of expression and the position of their expression domains along the anterior-posterior axis of the organism. This is described as the colinearity rule (Lewis, 1978) and it also coincides

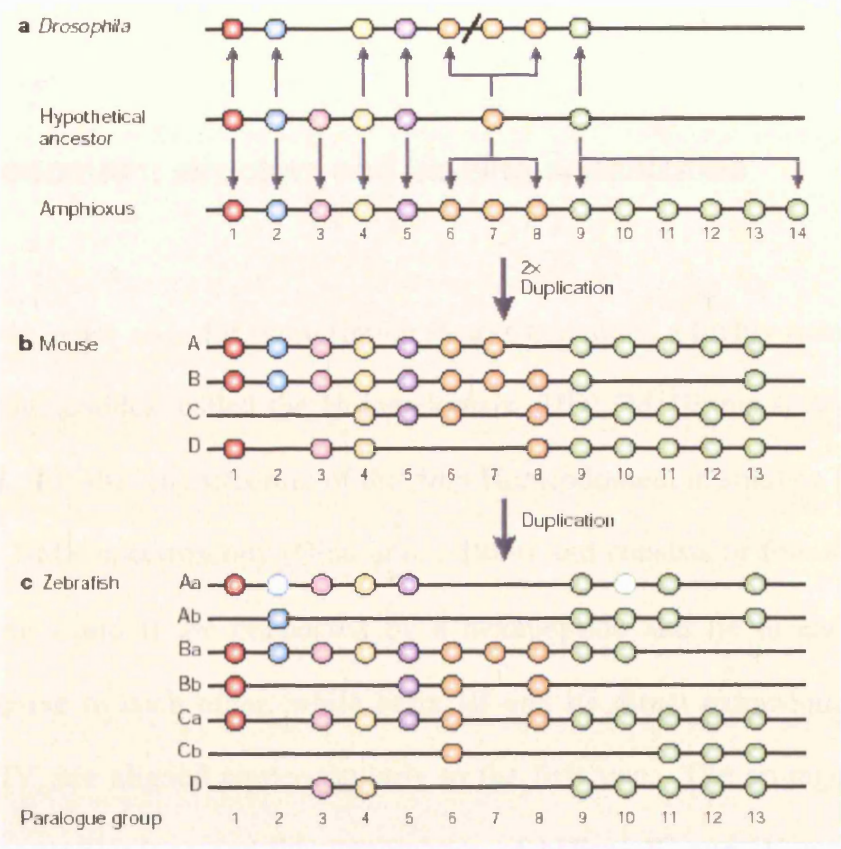


Figure 1.1: HOX cluster evolution. The evolution of the HOX cluster is characterised by multiple duplication events. *Drosophila* and other invertebrates have a single cluster with varying number of genes. Vertebrates have multiple HOX clusters as a result of tandem cluster duplications. These duplications might have just involved the clusters themselves, the chromosome on which they lie, or might have been whole genome duplication events as it has been suggested for zebrafish. Many genes have subsequently been lost, so that zebrafish has fewer than twice as many HOX genes as mouse. (Taken from Prince and Pickett, 2002)

with a functional hierarchy within the complex with the most posterior genes being dominant over their anterior counterparts.

1.1.1. Homeodomain: structure and binding specificities

Homeotic genes code for transcription factors that share a highly homologous region of 180 nucleotides, called the Homeodomain (HD) (McGinnis *et al.*, 1984a; McGinnis *et al.*, 1984b). The structure of the *Antp* Homeodomain in solution has been determined by NMR spectroscopy (Qian *et al.*, 1989) and consists of four α-helical regions: Helices I and II are connected by a hexapeptide and lie in antiparallel orientation relative to each other, while helix III and its direct extension, a more flexible helix IV, are aligned perpendicularly to the first two. The arrangement of helix II, the connecting loop and the combination of helices III and IV constitute a helix-turn-helix motif, very similar to what has been described for prokaryotic repressor proteins. This represents a striking evolutionary conservation in the three dimensional structure of the DNA binding motif from *E. coli* to *Drosophila* and -as it has been later confirmed- humans as well, even though there is hardly any sequence similarity at the amino acid level.

In vitro studies have shown that the purified Antp HD is sufficient to bind DNA in a sequence specific manner. Nevertheless, in contrast to the prokaryotic repressors that tend to form homodimers the Antp HD binds DNA *in vitro* as a monomer with a relatively high affinity, due to a low dissociation rate. Furthermore it has been suggested that the HD has a relatively high affinity for non-specific

interactions with DNA, which might enhance the overall stability of the complex (Affolter *et al.*, 1990).

NMR spectroscopy has also revealed the structure of the Antp Homeodomain complexed with DNA (Billeter and Wuthrich, 1993; Otting *et al.*, 1990; Qian *et al.*, 1992). The recognition domain consists of the helixes III and IV that interact with the major groove of the DNA, but other parts of the Homeodomain are also involved in direct protein-DNA interactions: the flexible N-terminal arm makes contacts with bases in the minor groove while the loop between helices I and II interacts with the DNA on the other side of the major groove. Moreover, as it has been shown for the Antp-DNA complex, several molecules of water are accommodated in a cavity that forms at the interface between the recognition helix and the DNA mediating hydrogen bonds between Homeodomain residues and the polar groups of the DNA (reviewed in Gehring et al., 1994a).

The binding specificities of the Homeodomain proteins have been extensively studied both *in vitro* and *in vivo*. Most of the DNA sequences that have been shown to interact with Homeodomains contain a ATTA (or TAAT, in the complementary strand) core (Gehring *et al.*, 1994b). This core comes into contact with residues of both the recognition helix (Ile47 and Asn51) and the flexible N-terminal arm, notably the very well conserved Arg3 and Arg5. The nature of the bases lying immediately adjacent to the ATTA core has also been shown to have an effect in determining binding specificities. Homeodomains with a Lysine (K) at position 50 (e.g. *bicoid*) seem to have a strong preference for a GG dinucleotide immediately upstream of the ATTA core (GGATTA) while those with a Glutamine (Q) at the same position (like *fushi tarazu, ftz*) bind with high affinity to a C[C/A]ATTA motif (Figure 1.2). A series

of elegant experiments with the Ftz protein have shown that changing the Ftz binding motif in the autoregulatory enhancer of the *ftz* region from CCATTA to GGATTA results in a reduction in Ftz's transcriptional activity *in vivo*, and that this effect can be specifically repressed by a DNA binding specificity mutant *ftz*Q50K (Schier and Gehring, 1992).

Interestingly the flexible N-terminal arm of the Homeodomain has also been shown to be functionally important. The recognition helices of the *Antp* and *Sex combs reduced (Scr)* gene products are identical. Changing four amino acids in the N-terminal flexible region of the Scr Homeodomain so that they match those in the Antp protein results in a hybrid protein that *in vivo* can mediate homeotic responses of the *Antp* type. i.e antenna to leg transformations (Furukubo-Tokunaga *et al.*, 1993).

But how can *in vivo* specificity be achieved when the level of conservation within the Homeodomain is so high? As already mentioned, residues outside the recognition helix have an effect on the sequence preference of the Homeodomain. In the case of Homeodomain proteins that also contain other DNA binding domains (see below), specificity is determined by the affinities of the two DNA binding domains for their respective DNA binding sites. Moreover the interaction with other protein cofactors (that may or may not bind DNA themselves) allows a further increase in the level of specificity.

Various experiments have shown that *in vitro* DNA binding specificity is not the sole determinant of the functional specificity exhibited by Homeodomain transcription factors.

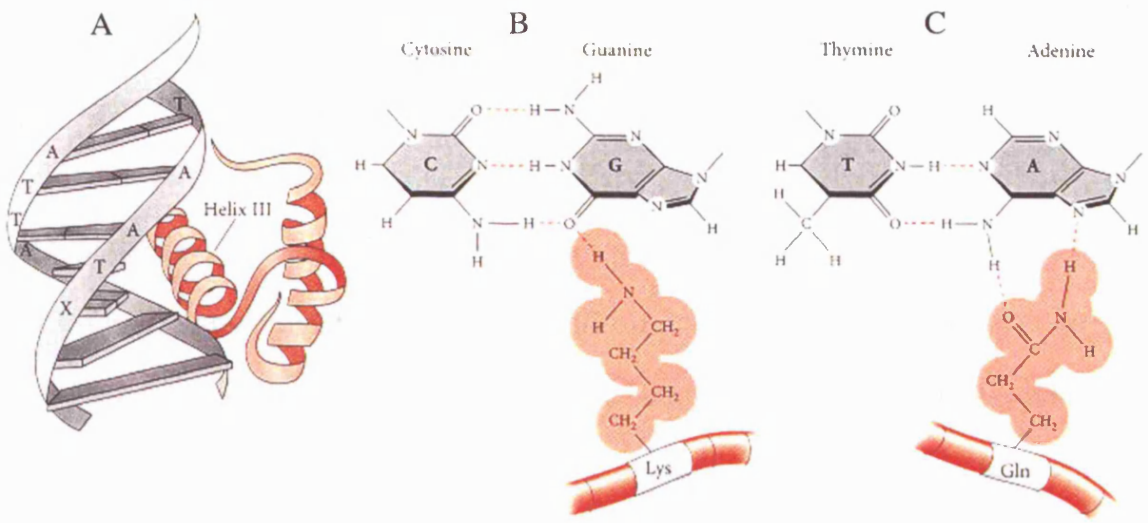


Figure 1.2: Homeodomain-DNA interactions. (A) Schematic representation of the Homeodomain-DNA complex. The 3rd helix of the Homeodomain interacts with bases in the major grove of the DNA. (B) Homeodomains with a Lysine at position 50 (e.g *bicoid*) have a preference for a GG dinucleotide immediately upstream of the ATTA motif (GGATTA). Lysine makes direct contacts with the proximal G via its side chain. (C) Homeodomains with a Glutamine at position 50 (e.g *antp*) prefer a C[C/A]ATTA motif. The interaction between the Glutamine side chain and the A preceding the ATTA motif is depicted. Modified from Gilbert S.F, Developmental Biology (6th edition, 2000)

The FtzQ50K mutant that *in vitro* exhibits the specificity of the Bcd protein cannot activate *in vivo bcd* target genes in the domains of *ftz* expression (Schier and Gehring, 1992). This can be explained by the hypothesis that Homeodomain transcription factors *in vivo* might not simply bind sites for which they have highest affinity *in vitro*. The degree of conservation of lower-affinity Ftz binding sites in *D. virilis* and *D. hydei* implies that they have some functional significance. An intriguing possibility is that lower-affinity binding sites for one HD transcription factor might act to prevent recognition of specific regulatory elements by other HD proteins expressed in the same cells and at the same time (Gehring *et al.*, 1994a).

1.1.2. Classification of Homeodomain sequences

1.1.2.1. The Complex and the Dispersed superclass

With respect to sequence similarity and chromosomal arrangement, homeotic genes have been classified into two major groups: classic homeotic gene complexes belong to the Complex superclass (Akam, 1989). All other HD containing genes belong to the Dispersed superclass, (Gehring *et al.*, 1994a) whose members have been largely dispersed in the genome and are more divergent in sequence. As a secondary criterion for the classification of different homeotic genes one can use the presence of other conserved sequence motifs outside the Homeodomain. With the exception of *Abd-B* all homeotic genes of the Complex superclass share a conserved "hexapeptide motif" IYPWMK that lies at the N-terminal end of the Homeodomain (Burglin, 1994). This motif is also found in some of the members of the Dispersed superclass implying

that these genes might have originally been part of a homeotic cluster and were subsequently translocated to other chromosomal positions (Gehring *et al.*, 1994a). The Dispersed superclass is more divergent on the basis of sequence and can be subdivided into smaller classes. Some of these have been determined based on the level of similarity within the Homeodomain, such as the *even skipped*, *engrailed*, *caudal* and *distalless* classes, whereas others are characterized by the presence of a second conserved motif besides the Homeodomain. In many cases this additional motif is a second DNA-binding domain that potentially confers greater binding specificities. For example, the *paired* class proteins contain a second DNA binding domain of 128aa known as the paired domain (Ton *et al.*, 1991), the *POU* class contains the 80aa long POU DNA binding domain (Herr *et al.*, 1988), while the *Zinc Finger* class contains some of the most unusual members with multiple zinc fingers and Homeodomains (Fortini *et al.*, 1991).

1.1.2.2. Atypical HD transcription factors: the TALE class

Finally there are some atypical groups that are characterized by insertions or deletions within the Homeodomain. Interestingly these insertions and deletions have a relatively minor effect in the three-dimensional structure of the Homeodomain. Members of the TALE class are characterized by the presence of three extra amino-acids between helices I and II (TALE stands for Three Amino-acid Loop Extension, reviewed in Burglin, 1997). There are four TALE subclasses in animals: TGIF (Bertolino *et al.*, 1995), MEIS (Steelman *et al.*, 1997), PBC (Flegel *et al.*, 1993; Nourse *et al.*, 1990) and Iroquois (IRO) (Leyns *et al.*, 1996).

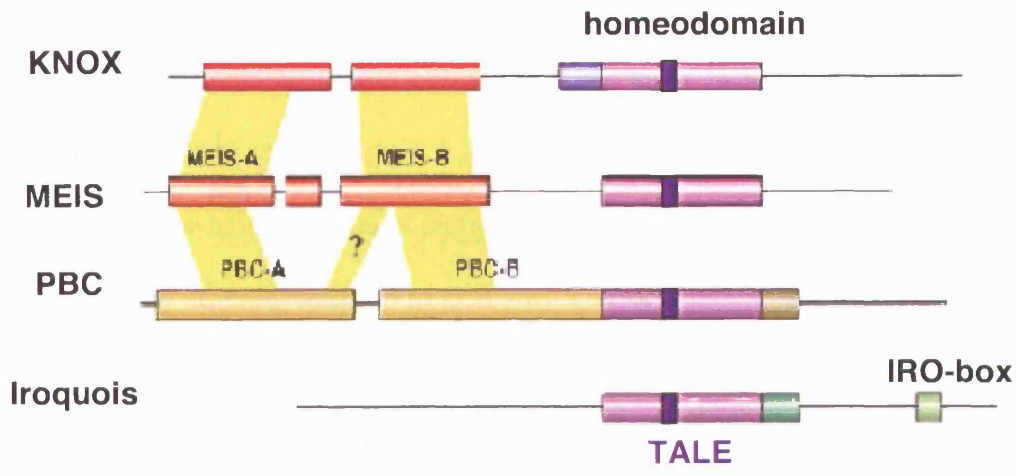


Figure 1.3: Alignment of the MEIS, PBC, Iroquois (animals) and KNOX (plants) families of TALE transcription factors. TALE family members have a Three Amino acid Loop Extension between helix 1 and helix 2 of their HD. Within the KNOX, MEIS and PBC families there are other domains of homology apart from the HD. The MEIS and KNOX domains share significant sequence similarity indicating that they may have derived from an ancient domain referred to as the MEINOX domain. Members of the Iroquois family do not share additional conserved motifs with the rest of the TALE members and are characterised by the presence of a short (13aa) domain of homology at the C-terminal end of the protein, which was named the IRO-domain. Taken from Bürglin, 1998.

There are also two in plants: KNOX (Vollbrecht *et al.*, 1991) and BEL (Reiser *et al.*, 1995) and two in fungi: M-ATYP (Astell *et al.*, 1981) and CUP (Burglin, 1997).

Most of the atypical classes are characterized by the presence of other conserved domains outside of the Homeodomain, such as the PBC, the MEIS, the KNOX and the IRO domain (Figure 1.3). A great level of similarity exists between the KNOX and the MEIS domains suggesting that they might have derived from a common ancestral domain, referred to as the MEINOX domain (Burglin, 1998).

Based on an interesting model the duplications that gave rise to the MEIS and PBC families in the animal branch of the evolutionary tree occurred at the same time as the initial duplications that led to the formation of the HOX cluster (Burglin, 1998). Members of the three families (MEIS, PBC and HOX) have been shown to interact both in flies and in vertebrates (see section 1.1.3) suggesting that they are part of a complex regulatory network acting to set up the anterior-posterior body axis (Figure 1.4).

Finally there are some more extreme cases of atypical HD proteins like the rat liver transcription factor LFB1 that contains 81 amino-acids in its Homeodomain rather than the typical 60. Strikingly these extra amino acids are accommodated in a flexible linker region between the first two α-helices without causing any major constraints on the overall conformation of the protein (Leiting *et al.*, 1993).

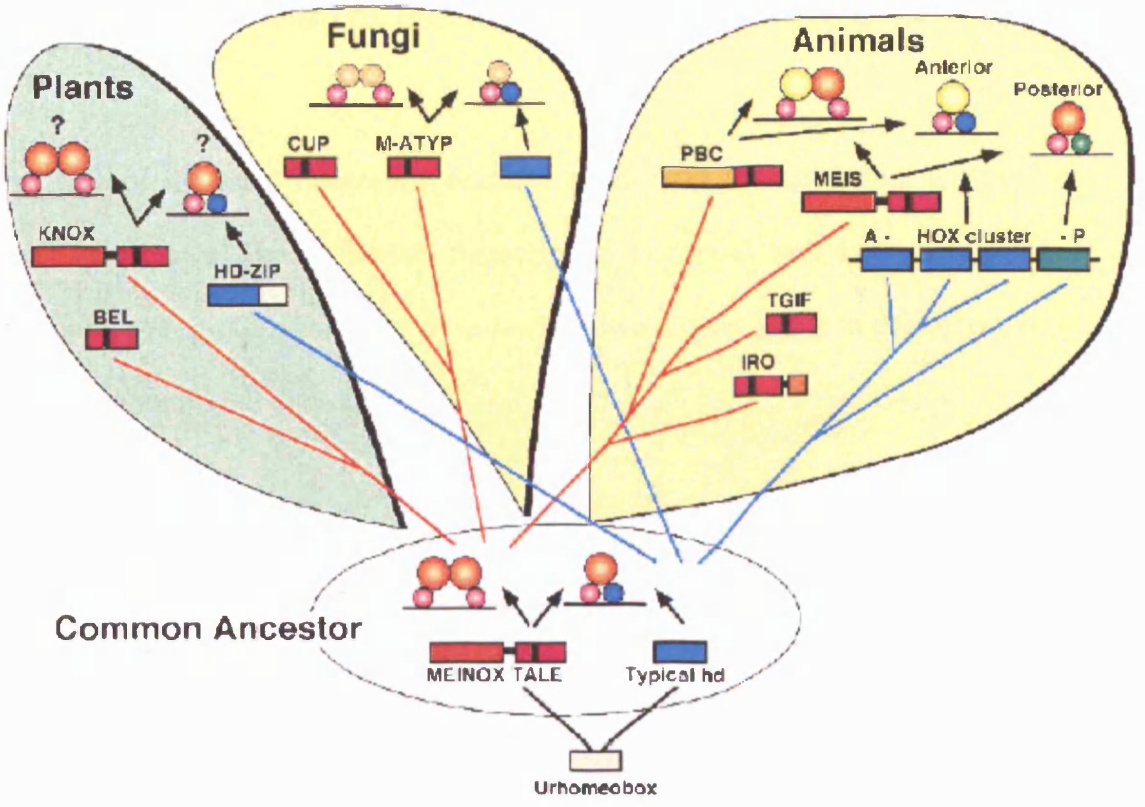


Figure 1.4: A model for the evolution of the TALE class of HD transcription factors. The common ancestor of fungi, plants and animals had one TALE gene that contained a MEINOX domain, possibly for protein-protein interactions. TALE proteins could form dimers (as seen today with PBC/MEIS members) and interact with the ancestral HOX gene product. In animals the ancestral TALE gene has given rise to the TGIF, IRO, PBC MEIS families. Evolution of the TALE class occurred in parallel with the evolution of the HOX cluster generating a complex regulatory network that involves interactions amongst members of all these groups to regulate targets genes and specify the anterior/posterior axis (taken from Bürglin, 1998)

1.1.3. Members of the TALE class exhibit cooperative DNA binding

One of the most interesting features of the TALE class is that its members interact with typical Homeodomain transcription factors as well as with each other. Interactions with protein cofactors have been shown to play a role in enhancing target specificity in numerous occasions.

One of the first documented cases was for the yeast atypical Homeodomain protein MATα2, a regulator of the mating type, that recognizes different target sequences depending on the protein cofactor it associates with (Goutte and Johnson, 1993; Vershon and Johnson, 1993).

The most extensive studies have been performed on the interaction between members of the PBC family and members of the homeotic gene complex (HOX). These studies were triggered by the observation that in *Drosophila,* mutations in the gene *extradenticle (exd),* which is part of the PBC family, can cause homeotic transformations similar to those caused by *Hox* gene mutations (Peifer and Wieschaus, 1990). *exd* is not itself a regulator of homeotic gene expression nor does its activity depend on regulation by homeotic gene products. It was therefore suggested that *exd* and *Hox* gene products may act in parallel to regulate expression of target genes. This was confirmed by crystallographic studies showing that Exd and HOX proteins can cooperatively bind to DNA (Figure 1.5). This interaction is also conserved in mammals where members of the PBC family PBX1, 2 and 3 also bind DNA in a complex with mammalian HOX proteins (reviewed in Mann and Chan, 1996).

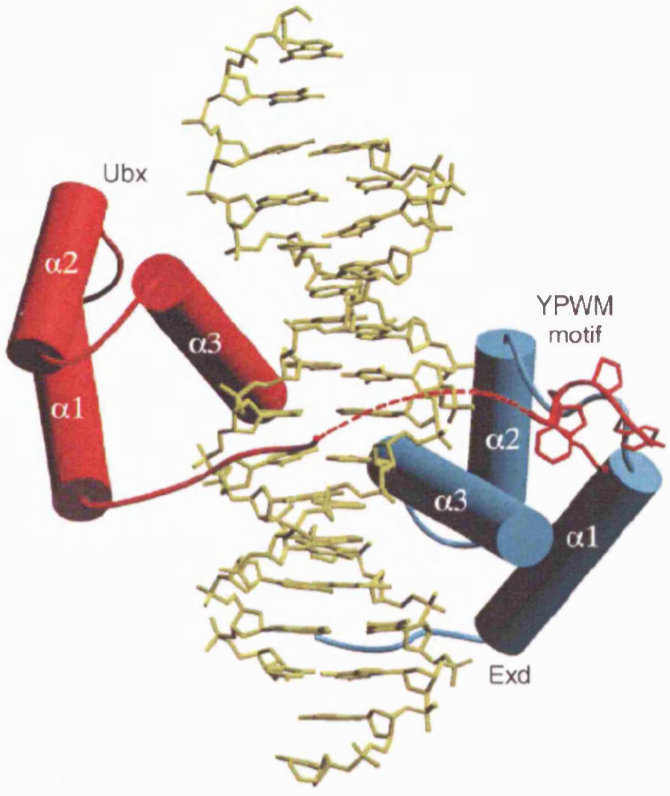


Figure 1.5: Structure of the Ubx-Exd complex on the DNA. Ubx (red) and Exd (blue) Homeodomains approach DNA from opposite sides and bind in a tandem conformation. The YPWM motif at the N-terminal of Ubx protein comes in close proximity and interacts with a hydrophobic pocket on the surface of Exd protein. Note than in both cases the 3rd helix of the Homeodomain is positioned within the major groove of the DNA. Taken from Passner *et al*, 1999

This interaction is dependent on the YPWM motif that lies at the N-terminal end of the HOX Homeodomain (Johnson *et al.*, 1995) and residues within and C-terminal to the PBC Homeodomain (Lu and Kamps, 1996).

Binding site selection assays have identified a bipartite sequence (5'-ATGATTNATNN-3') consisting of two defined half sites for each of the two monomers: the 5' ATGAT motif is bound by the PBX protein while the more variable 3'half (TNATNN) is recognized by the HOX component of the heterodimer. Interestingly, different HOX proteins acquire different binding specificities for DNA upon heterodimerisation with PBX. This is believed to occur through a PBX-induced conformational change in the HOX HD N-terminal arm thereby providing a mechanism to increase functional specificity within the HOX locus (Chang *et al.*, 1996). In *Drosophila,* Exd interacts with the product of the *Hox* gene *labial* to autoregulate *labial* expression *in vivo*. In the absence of Exd, Labial protein cannot bind the DNA region that mediates this transcriptional control. Mutations in its YPWM motif allow Labial protein to overcome the requirement for Exd to regulate its own transcription, suggesting that Exd binding may relieve an inhibitory function of the YPWM motif therefore changing the *in vivo* specificities of the Labial protein (Chan *et al.*, 1996).

Members of the PBC family, in addition to interacting with HOX proteins, have also been shown to interact with other TALE proteins, namely with members of the MEIS family. The observation that led to the characterisation of the PBC-MEIS interaction was that mutants of the *Drosophila* MEIS homologue *homothorax (hth)* give similar phenotypes to *exd* suggesting that they might be involved in the same process. Furthermore *hth* has been shown to be required for *exd* function (Rieckhof *et*

al., 1997). Further studies have revealed that Hth acts to prevent Exd export from the nucleus, thereby maintaining its nuclear localisation (Berthelsen *et al.*, 1999). This was the first reported case of a HD transcription factor acting on the subcellular localisation of another HD protein showing that transcription factors may play additional, quite unexpected roles within a cell.

Another model for Exd/PBX mode of function suggests that these proteins are required for changing the function of HOX transcription factors from repressors to activators. According to this model Exd/PBX is a generic cofactor for HOX proteins. Part of their cofactor functions may be to shift their partners binding specificities towards a composite HOX/PBX binding site but according to Pinsonneault and co-workers this is neither a primary nor an essential function. These have shown that in the case of the autoregulation of the *Hox* gene *Deformed* (*Dfd*) the interaction between Exd and Dfd enables *Dfd* to act as an activator whereas in the absence of Exd it would act as a repressor. (Pinsonneault *et al.*, 1997).

1.2. The Iroquois family of transcription factors

The *Iroquois* (Iro/Irx) family is a highly conserved class of TALE genes with members in organisms as evolutionarily distant as sponges, nematodes, *Drosophila* and humans. They were first identified in *Drosophila* through mutations that suppressed the lateral bristles of the dorsal mesothorax (notum) of the fly, leaving only a median stripe of hairs (Leyns *et al.*, 1996). This phenotype was reminiscent of a Mohawk, the hairstyle of the American Indian tribe Iroquois, hence the name of the

locus. The *Iroquois* genes were shown to act as prepattern factors necessary for the expression of proneural genes in *Drosophila* (Bosse *et al.*, 2000; Gomez-Skarmeta *et al.*, 1996; Gomez-Skarmeta and Modolell, 1996; Kehl *et al.*, 1998; Leyns *et al.*, 1996). In flies there are three *Iroquois* genes named *araucan (ara), caupolican (caup)* -for an Iroquois tribe and its leader respectively- and *mirror (mirr).* The latter was the last one to be identified (McNeill *et al.*, 1997) and has been named for its role in establishing a mirror image pattern in the fly eye.

Comparison of the worm, fly and human Iroquois sequences revealed that similarity is mainly restricted to the Homeodomain (Burglin, 1997). The Iroquois Homeodomain is indeed very well conserved amongst homologues from all species. *Drosophila* Mirror shares a 95% identity within the Homeodomain with mouse Irx4. The homology outside the HD however is not very high: Ara and Caup share a 41% of overall homology, while Mirror is more divergent and shares a 27% identity with Ara.

Iroquois members also share a novel motif of homology that lies in the C-terminal part of the molecule and has only been found in members of this family (McNeill *et al.*, 1997). This domain has therefore been named the Iro-box (Burglin, 1997). This short motif has been thought to mediate protein-protein interactions but to date there is no evidence for homo- or heterotypic interactions mediated by this motif.

Other domains of homology include a EGF-like motif at the N-terminal end of the molecule that has been suggested to be involved in protein-protein interactions (Bosse *et al.*, 2000) and a region of highly acidic residues downstream of the Homeodomain reminiscent of a transcription activation domain (Burglin, 1997).

1.2.1. Genomic organization of the *Iroquois* genes

With the exception of *C.elegans* where there is only one *Iroquois* gene, in all other organisms, where they have been identified, *Iroquois* genes form clusters: In *Drosophila*, *ara, caup* and *mirror* are clustered within 130kb of genomic DNA on the left arm of chromosome 3. The distance between *ara* and *caup* is less than 25kb while *mirror* is located ~70kb downstream of *caup* (Figure 1.6). *ara* and *caup* have almost identical expression patterns, suggesting that their expression might be under the control of shared regulatory elements. They are also thought to be functionally redundant (Cavodeassi *et al.*, 2001; Gomez-Skarmeta *et al.*, 1996). *mirror* has a more divergent expression pattern with some overlap with the other two in the eye disc, the lateral epidermis, the proventriculus, the brain and a region of the wing disc (Gomez-Skarmeta *et al.*, 1996; McNeill *et al.*, 1997). *mirror* is additionally expressed during early embryonic stages, its expression starts as early as stage 5 (130-180min after egg laying) and is important for the development of the embryonic CNS, whereas *ara* and *caup* are not expressed in the CNS (Mohns, 2003).

The genomic organization of the *Iroquois* cluster in *Drosophila* seems to be conserved in vertebrates. In mouse and humans there are 6 genes, arranged in two clusters of three: *Irx1*, *Irx2* and *Irx4* are on chromosome 13 and form the *IrxA* cluster, while their respective paralogues *Irx3*, *Irx5* and *Irx6*, are on chromosome 8 and form the *IrxB* cluster (Peters *et al.*, 2000 and Figure 1.6).

Several pieces of data suggest that the two vertebrate clusters have derived as a result of a chromosomal duplication of an ancestral cluster: First, each member of a cluster is most similar to the gene found in the same position of the other cluster.

Second, similar to what is seen in *Drosophila*, the pattern of expression of the first two genes within a cluster is highly similar, while the third one is slightly divergent (Houweling *et al.*, 2001). Third, at least in humans and zebrafish, sequences flanking the *Iroquois* clusters have conserved genes. Comparison however of the *Drosophila* and vertebrate proteins has shown that the latter are more similar to one another than to their fly counterparts (Peters *et al.*, 2000). This implies that the duplications that resulted in the generation of the *Iroquois* clusters in flies and vertebrates may have occurred independently in the ancestors of the insect and the vertebrate lineage. What is intriguing though is that in both cases clusters consist of 3 genes each.

As mentioned above the clustering of *ara* and *caup* in *Drosophila* allows them to be controlled by common regulatory elements, as evinced by their almost identical expression patterns. It is possible that these common elements in some cases extend to the control of *mirror* expression. As judged by their overlapping expression patterns vertebrate *Iroquois* might also share regulatory elements. This sharing of regulatory elements amongst members of the same cluster could be a plausible explanation as for why this genomic organisation has been maintained during evolution.

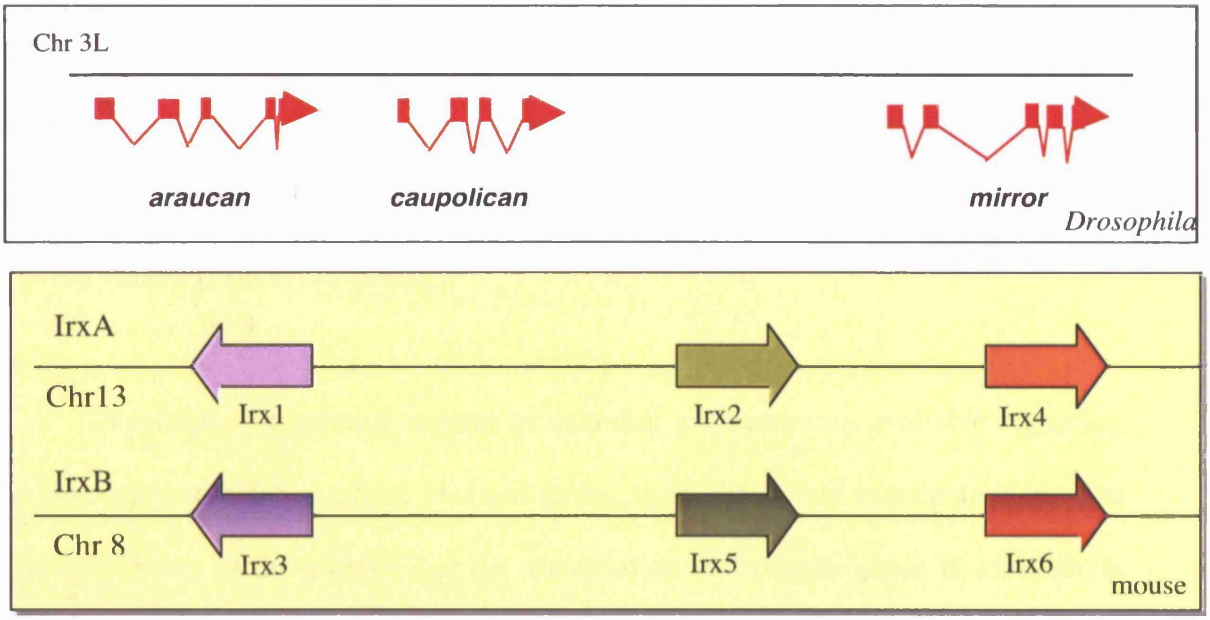


Figure 1.6: Genomic organisation of *Iroquois* genes in *Drosophila* and vertebrates. In flies there is a single cluster that consists of three genes on the left arm of chromosome 3. In mouse there are two clusters located on different chromosomes (8 and 13). Pairs of paralogous genes (*irx1/irx3, irx2/irx5* and *irx4/irx6*) share a greater similarity with one another than with other members of their own cluster and are transcribed in the same orientation. Note that orientation of transcription differs between the *Drosophila* and vertebrate clusters.

In zebrafish four *Iroquois* genes have so far been identified, their genomic organisation however is still unknown. *ziro3* and *5* map to a single linkage group, suggesting that, like mouse *Irx3* and *5,* they are members of the same cluster. *ziro 1* and *7,* however, map in different linkage groups. To date it is not clear if this is due to the presence of an additional cluster in zebrafish (consistent with the extra gene duplication that took place during teleost evolution) or if it reflects a break up of an existing cluster (Itoh *et al.*, 2002).

Despite the increasing amount of data that are becoming available regarding the genomic organisation of the *Iroquois* genes, many important aspects remain to be clarified. Study of the number and the organisation of *Iroquois* genes in more phyla should help determine the timing of the various duplication events that gave rise to the full complement of *Iroquois* gene in the animal kingdom.

1.3. The role of the Iroquois genes in development

A general feature of *Iroquois* genes in all species is their function in specifying borders. In a rather simplified generalisation, one can say that they act at early stages of development to define large territories while at later stages they are responsible for specifying smaller domains within these territories (reviewed in Cavodeassi *et al.*, 2001). In the following sections I will present a summary of the current knowledge of *Iroquois* function in vertebrates and in *Drosophila.*

1.3.1. Iroquois function in vertebrate development

1.3.1.1. Formation of the neural plate

The first vertebrate *Iroquois* genes identified were the *Xenopus* homologues (*Xiro 1, 2* and *3*), which are essential for neural development. *Xiro1* and *Xiro2* are expressed at the beginning of gastrulation in the dorsal ectoderm and are essential for the specification of the neural plate. Injection of mRNAs coding for wild type Xiro proteins has shown that their overexpression results in expansion of the neural plate and reduction of the neural crest (Bellefroid *et al.*, 1998; Gomez-Skarmeta *et al.*, 2001; Gomez-Skarmeta *et al.*, 1998). Interfering with *Xiro*1 function by injecting a dominant negative construct suppresses neural differentiation, promoting instead the epidermal fate (Figure 1.7).

It had been previously shown that formation of the neural plate in *Xenopus* depends on suppression of *Bmp-4*, which, when active, promotes the epidermal fate. This downregulation is achieved through two separate mechanisms: Firstly it depends on the presence of neural inducers such as Noggin, Chordin and others that directly bind to Bmp-4 and prevent its interaction with the Bmp-4 receptor. In addition, the Wnt pathway represses expression of Bmp-4 in the prospective neuroectoderm. *Xiro1* and most probably *Xiro2* are activated by the Wnt pathway and act to repress Bmp-4 (Gomez-Skarmeta *et al.*, 2001). This repression may be direct, as it has been shown that a partial GST-Xiro1 construct (lacking 41 residues at the N-terminal) can bind to a fragment of the *Bmp-4* promoter. *Xiro1* also represses *Bmp-4* during dorsal mesoderm formation (Glavic *et al.*, 2001).

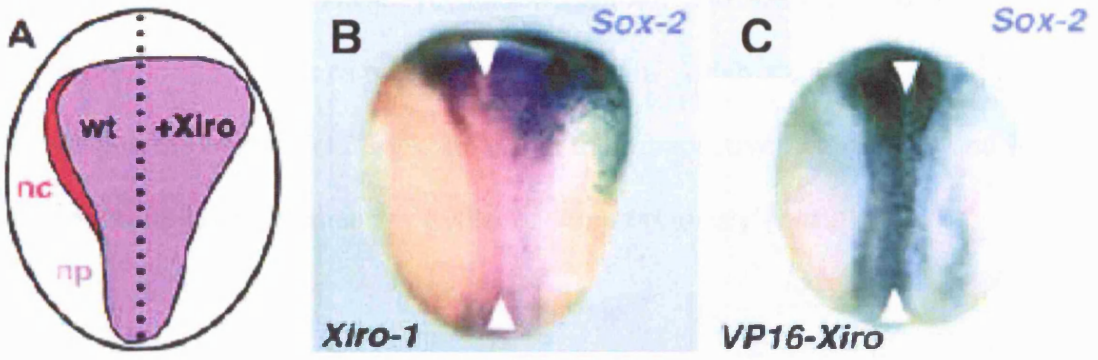


Figure 1.7: Iroquois are important for neural plate specification. (A) A schematic representation of a *Xenopus* embryo at the neurula stage injected at the right hand side with Xiro mRNA. The neural plate at the side of the injection expands at the expense of the neural crest. (B) Embryos injected with Xiro1 and β-gal mRNA. Staining with the Sox2 marker shows the extent of the neural plate expansion at the side of the injection. The blue-green staining (Xgal) reveals the injected side. (C) Injection of a Xiro1 fusion to the transcriptional activator VP16 results in almost complete suppression of neural differentiation at the injected side, indicating that the VP16-Xiro construct acts as a dominant negative and that Xiro1 acts as a repressor. Taken from Gómez-Skarmeta *et al*, 2001.

Xiro3 has also been shown to have neuralising properties although it is expressed later in development (Bellefroid *et al* 1998). Interestingly the zebrafish *ziro3* also has been shown to be downstream of Wnt signalling and to downregulate *Bmp-4* in the context of organiser formation (Kudoh and Dawid, 2001). The regulation of *Bmp-4* may prove to be a general feature of vertebrate *Iroquois* as it has been reported that the chick *cIrx2* is expressed in the prospective neural plate and its pattern of expression is complementary to that of *Bmp-4* (Goriely *et al.*, 1999).

Recently, *Iroquois* genes in *Xenopus* and zebrafish have been implicated in the specification of the neural crest (Glavic *et al.*, 2004b) and the preplacodal region, a domain of thickened epidermis at the border between the neural plate and the epithelium that contributes to the formation of nose, eyes, ears, lateral line and cranial sensory ganglia (Glavic *et al.*, 2004a).

1.3.1.2. Subdivision of the neural plate

After the neural plate is specified, it folds to form the neural tube, which then becomes subdivided into different territories (Figure 1.8). The Dorsal/Ventral subdivision is a result of the coordinated action of signalling molecules emanating from the dorsal roof and the ventral floor plate. A gradient of Sonic hedgehog (Shh) acts to set up the domain of expression of two sets of transcription factors. The first set (class-I genes) are expressed in nested domains from dorsal to ventral regions as a result of repression by different concentrations of the gradient. Class-II genes are activated by the Shh gradient and are consequently expressed in nested domains from ventral to dorsal regions (reviewed in Gomez-Skarmeta and Modolell, 2002).

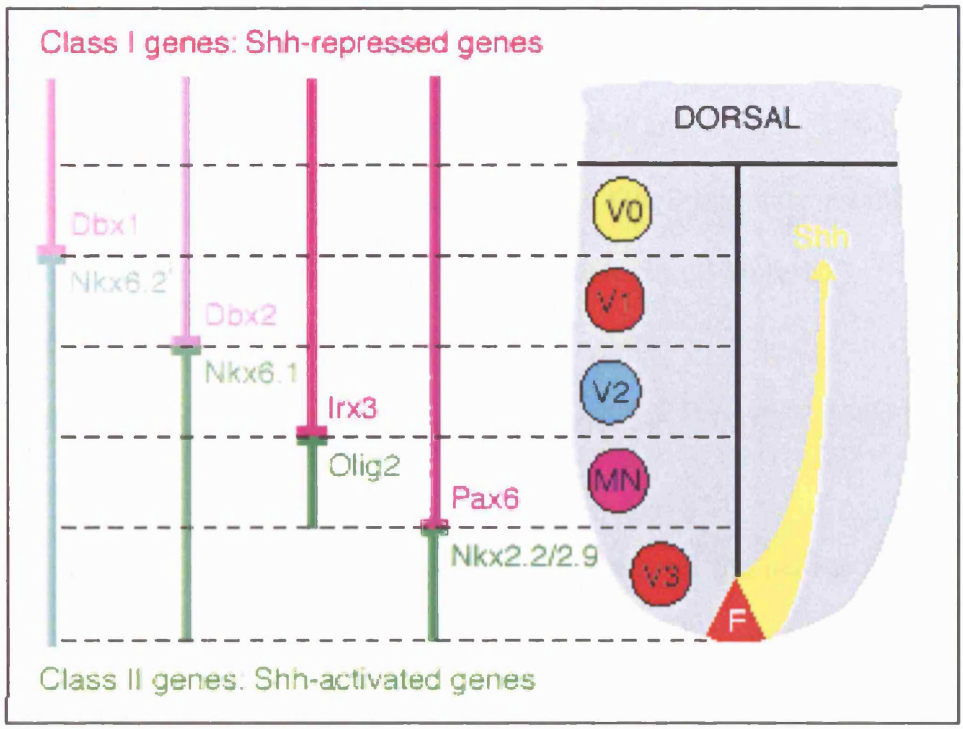


Figure 1.8: Iroquois proteins are involved in the subdivision of the ventral spinal cord. Cartoon of a transverse section showing that mutual repression between pairs of transcription factors subdivides the spinal cord into five regions, from which different types of neurons will form (V, ventral, MN, motorneurons). Class-I genes are repressed by various levels of Sonic hedgehog (Shh), while class-II genes are activated by different levels of the Shh gradient. *Irx3* is one of the class-I genes and is repressed by Shh. *Irx3* together with the bHLH transcription factor Olig2 act to specify the border between the motor neuron (MN) and the second ventral region (V2). Taken from Gómez-Skarmeta and Modolell, 2002.

A mutual repression between pairs of one class-I and one class-II gene contributes to the sharpening of their borders of expression resulting in the establishment of 5 distinct zones, from which different subtypes of neurons will arise. *Irx3* is a class-I gene. It is repressed by the Shh gradient and by the class-II HLH transcription factor Olig2. *Irx3* in its turn acts to repress *Olig2* therefore establishing the border between the second Ventral and Motor Neuron region (Figure 1.8).

1.3.1.3. *Iroquois* function in the vertebrate brain

Patterning of the vertebrate brain involves a division of the neuroepithelium into the three brain areas: forebrain, midbrain and hindbrain, which are in turn subdivided into subdomains. For example the hindbrain divides into a series of rhombomeres and the rhombic lip at the dorsal edge of the hindbrain gives rise to the cerebellum. Similar to the subdivision of the neural plate, mutual repression between pairs of HD transcription factors acts to subdivide the brain along the Anterior/Posterior axis (Figure 1.9). In the chick, Iroquois proteins are involved in the subdivision of the forebrain. Irx3 and the HD transcription factor Six3 mutually repress each other and subdivide the forebrain into anterior and posterior domains providing competence for the spatially restricted expression of neural genes in response to diffusible signals such as Shh and FGF8 (Kobayashi *et al.*, 2002). Furthermore, overexpression of Irx3 causes a forebrain to midbrain transformation, implying that Irx3 is also involved in midbrain specification.

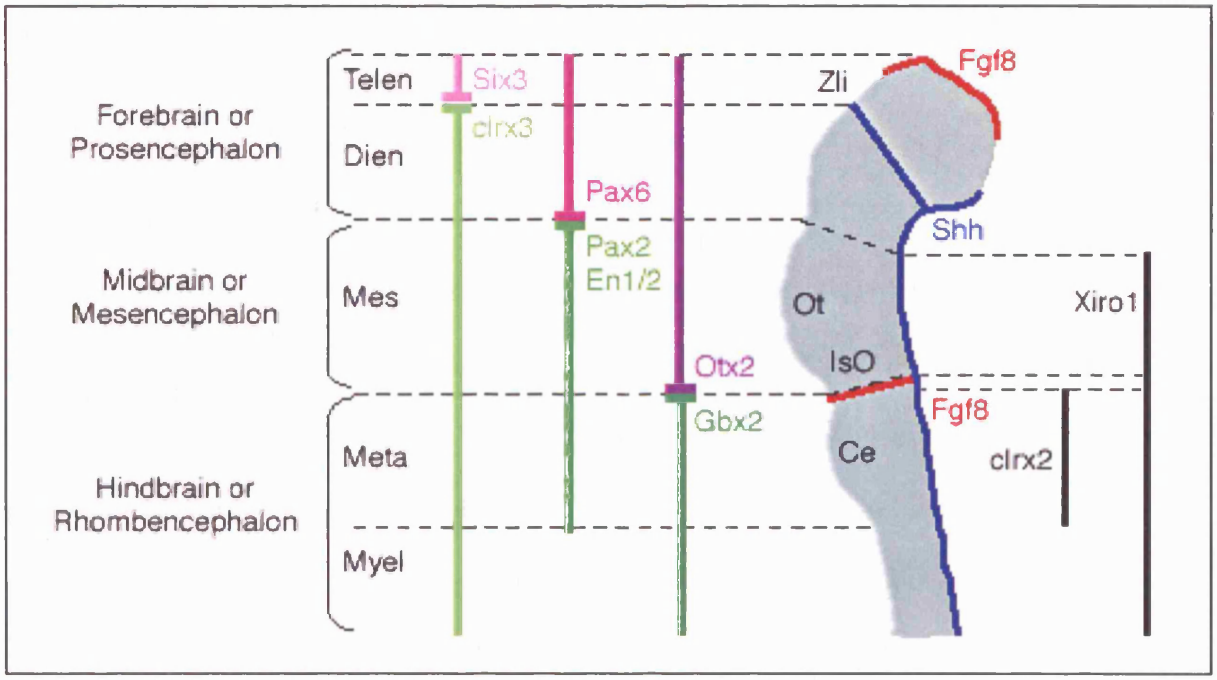


Figure 1.9: Iroquois proteins function in the subdivision of the vertebrate brain. Similar to what was described for the ventral spinal cord, the developing brain is subdivided into regions due to the mutual repression between pairs of transcription factors. *Iroquois* genes are expressed in distinct domains. Chick *Irx2* is expressed in and required for specification of the cerebellum (Ce) region. In *Xenopus* the anterior border of *Xiro1* activity coincides with the forebrain-midbrain border. Taken from Gómez-Skarmeta and Modolell, 2002.

In zebrafish two *Iroquois* family members, *ziro1* and *ziro7* are essential for formation of the midbrain/hindbrain boundary and establishment of the isthmic organizer, a signalling centre important for the development of the midbrain and anterior hindbrain. Their activity is dependent on Wnt signalling and they have been shown to induce expression of the proneural gene, *neurogenin 1* (Itoh *et al.*, 2002).

Another member of the *Iroquois* family, *Irx2* is strongly expressed in the prospective cerebellum territory and has been shown to be important for cerebellum formation (Matsumoto *et al.*, 2004). In this context and in the presence of FGF signalling Irx2 acts as a transcriptional activator, while in the absence of the FGF8 signal Irx2 acts as a repressor. This activity switch is obtained by phosphorylation of residues in the N-terminal end of the molecule by the MAP kinase cascade. Phosphorylation by MAP kinase disables a repressor function attributed to the C-terminal end of the molecule. In summary FGF8/MAP kinase signalling acts as a molecular switch that modifies the transcriptional activity of Irx2. It will be very interesting to find out how this translates in terms of downstream targets and binding specificities, if, for instance, the switch from activator to repressor is accompanied by a switch in the DNA binding preferences and /or associated cofactors.

Mirror, Ara and Caup also have sites for potential MAPK phosphorylation in their N-terminal region, raising the possibility that this sort of regulation may also occur in flies. So far, however, there is no direct evidence linking FGF signalling with the *Iroquois* genes in *Drosophila.*

1.3.1.4. *Iroquois* and axonal pathfinding

During CNS development axons have to navigate long distances following specific paths and establishing neuronal connections. Several molecules have been implicated in the process of axon guidance acting either as attractants or as repellents. Like many guidance cues the Slit family of proteins can act both positively and negatively on axon pathfinding. It was recently reported that chick Irx4 acts to repress Slit1 expression in the retina and that this regulation is important for correct guiding of the retinal axons inside the optical fibre layer (Jin *et al.*, 2003). It is possible that other Iroquois proteins are involved in the regulation of Slit and that additional input mechanisms are required to pattern the trajectory of the axons with the retina. Interestingly *mirror* has also been shown to act on axonal pathfinding in the *Drosophila* CNS (Mohns *et al*, manuscript in preparation) implying that this may be a universal function of Iroquois proteins.

1.3.1.5. *Iroquois* and patterning of the vertebrate heart

The vertebrate heart develops from a single tubular structure, the heart tube, through a series of morphogenetic movements involving looping and curving leading to its subdivision into ventricular and atrial compartments. Atrial and ventricular chambers are highly specialised to enable them to perform their distinct roles in circulating blood. Specification of these compartments depends on the activity of transcription factors that control the expression of chamber specific genes. Irx4 has a conserved role in heart development in mice and chicks. In the mouse Irx4 is expressed early in the linear heart tube and its expression is restricted to the ventricular compartment (Bruneau *et al.*, 2000). In the chick heart Irx4 is also restricted to the

ventricles during all stages of cardiac development and its expression persists up to adulthood (Bruneau *et al.*, 2001). Four other *Iroquois* genes have been shown to be expressed in the developing mouse heart (Christoffels *et al.*, 2000), with greatly overlapping patterns (Figure 1.10).

Irx4 protein regulates chamber specific expression of the myosin isoforms by activating the ventricle-specific isoform (VMHC1) and at the same time repressing the atrial-specific isoform (AMHC1) (Bao *et al.*, 1999). Repression of the MyHC3 (myosin heavy chain) gene in quail by Irx4 requires a Vitamin D Response Element (VDRE) located 5' of the gene (Wang *et al.*, 2001). For this purpose Irx4 associates with the Retinoic X receptor (RXR), which as part of a RXR/VDR (Vitamin D Receptor) heterodimer mediates the VDRE-dependent transcriptional repression. Irx4 itself does not bind to the VDRE and residues within its N-terminal are required for its inhibitory action.

Mouse knockouts for the *Irx4* gene are viable but develop cardiomyopathy accompanied by aberrant gene expression patterns including ventricular expression of atrial-specific genes (Bruneau *et al.*, 2001). These results indicate that *Irx4* is not essential for ventricular chamber formation but is probably required for some aspects of the ventricle-specific gene expression pattern. The fact that several *Iroquois* genes are expressed in the heart and in highly overlapping patterns suggests that their generic role during cardiac development may be to refine the spatial regulation of chamber specific gene expression.

Figure 1.10: Iroquois expression during mouse heart development. Expression of the 5 *Iroquois* genes at stages E8-9.5 and E9.5-12 is shown. The levels of Irx mRNAs have been measured by *in situ* hybridisation and the grey and black colours correspond to low and high levels respectively. *Iroquois* expression is mostly restricted to the ventricular part. *irx3* and *irx4* have show stronger and wider expression patterns while the other three genes are not so highly expressed. AVC, atrioventricular canal, LA, left atrium, RA, right atrium, LV, left ventricle, RA, right ventricle, EA embryonic atrium, EV embryonic ventricle. Image taken from Christoffels *et al*, 2000.

1.3.2. Iroquois function in *Drosophila*

mirror and the other two *Iroquois* genes are expressed in various tissues and different developmental stages during *Drosophila* development. As discussed above *ara* and *caup* share a very similar expression pattern while expression of *mirror* is more divergent. In the following sections I will present a summary of the available information on the role of *mirror* and the other two *Drosophila Iroquois,* starting from their function during early developmental stages (oocyte and embryo) and then in the patterning of adult structures such as the wing, the notum and the eye.

1.3.2.1 mirror acts to establish the dorsal-ventral axis of the oocyte

Oogenesis in *Drosophila* starts in the germarium, the anterior-most structure of the ovariole with a germ-line cell initiating a series of 4 cell divisions that result in the generation of 16 cells. One of these will become the oocyte while the other 15 will be made into nurse cells responsible for providing macromolecules and organelles to the developing oocyte. All 16 cells are encapsulated by epithelial follicle cells. Follicle cells also derive from the gonads but are of somatic rather than germ-line origin.

Signalling emanating from subgroups of follicle cells is important for establishment of the anterior-posterior (A/P) and dorsal-ventral (D/V) axis of the egg and ultimately the embryo (reviewed in Riechmann and Ephrussi, 2001). In both cases signalling involves the EGFR pathway and is initiated by the ligand Gurken. The first axis to be established is the A/P axis, as evinced by the posterior positioning of the

oocyte by the beginning of stage 7. The D/V axis of the eggshell and the embryo is set upon activation of the Gurken/EGFR signalling pathway in the anterior dorsal follicle cells. Cells that receive the signal differentiate into midline or dorsal follicle cells, while the rest of the follicle cells acquire a ventral fate. This is the result of the activation of one set of genes and at the same time the repression of another set of genes in the anterior dorsal cells resulting in their restricted expression in the ventral cells. This is essential for the initiation of a complex signalling cascade that leads to the generation of a gradient of the protein Dorsal, which defines the embryonic D/V axis.

In the initial expression pattern study (McNeill *et al.*, 1997) *mirror* was described to be expressed in the dorsal anterior follicle cells. In a more in-depth analysis of its function during oogenesis Mirror has been shown to be important both at early and late stages of oocyte formation. Early on, it is required for the encapsulation of the prospective oocyte in the germarium by the follicle cells. At stage 6 egg chambers *mirror* is expressed in the lateral follicle cells. By stage 10 *mirror* is restricted to the anterior dorsal follicle cells and the centripetally migrating cells. At the same stages the gene *fringe* is expressed in non-overlapping, complementary patterns (Jordan *et al.*, 2000 and Figure 1.11).

In later stages *mirror* acts on the establishment of the D/V axis of the eggshell and subsequently of the embryo (Jordan *et al.*, 2000). *mirror* is activated by the Gurken/EGFR signalling pathway in the dorsal anterior follicle cells (Jordan *et al.*, 2000; Zhao *et al.*, 2000) and it has been shown to repress expression of *fringe* in ventral and posterior follicle cells (Jordan *et al.*, 2000).

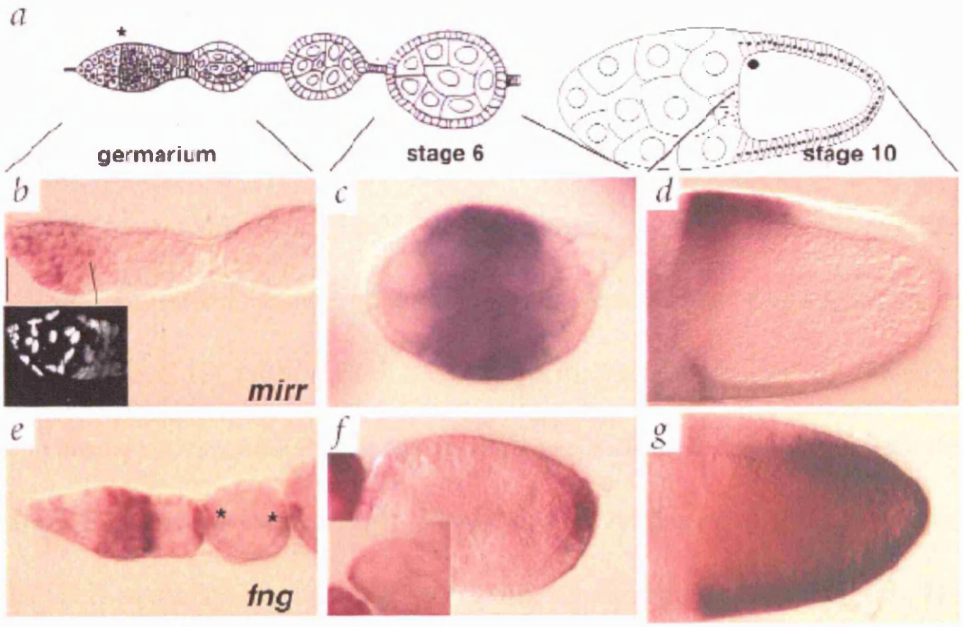


Figure 1.11: The *mirror* and *fringe* expression patterns are complementary during *Drosophila* oogoenesis. (a) Schematic of various stages of oogenesis (for description see text). (b-d) *mirror* expression in the germarium, stage 6 and stage 10 as detected by *in situ* hybridisation (inset in b: antibody staining for Mirror protein confirming the *in situ* results). (e-g) *fringe* expression at the same stages. Patterns are non-overlapping. In the germarium *mirr* is expressed in the anterior most cells while *fringe* in the posterior regions. At stage 6 *mirr* expression is restricted to the lateral and *fringe* to the terminal follicle cells of the egg chamber. Finally at stage 10 *mirr* is expressed in the anterior-dorsal and *fringe* in the remaining follicle cells. Taken from Jordan *et al*, 2000.

In a number of systems, regulation of *fringe* by a homeobox protein acts to define the borders of Notch expression (Irvine, 1999). In agreement with these observations Jordan and co-workers have shown that *mirror* and *Notch (N)* mutants have similar phenotypes during oogenesis : Eggs laid by N^{ts} as well as *mirror* loss-of-function females display a ventralisation of the eggshell as shown by the loss of dorsal structures such as the dorsal appendages. Furthermore, similar to *mirror* loss-of-function, *Notch* mutations cause an expansion in the ventral expression of *pipe,* another gene required for embryonic D/V axis formation (Jordan *et al.*, 2000). Pipe initiates a proteolytic cascade that leads to the production of the ligand Spätzle that binds to the Toll receptor (reviewed in Morisato and Anderson, 1995) Activation of the Spätzle/ Toll pathway leads to the generation of a gradient of the protein Dorsal, which is essential for establishment of the D/V axis of the embryo (Stathopoulos and Levine, 2002a).

According to the current working model, the Mirror-Fringe border activates Notch in a restricted domain to produce a still unidentified morphogen that acts at a distance to repress the expression of *pipe*. The *Bmp4* homologue *dpp* is a candidate for the secreted diffusible molecule that represses *pipe*. Dpp is expressed in a domain adjacent to the Notch expression stripe and in a Notch dependent manner (Jordan *et al.*, 2000). Furthermore mutations in the downstream effectors of the *dpp* pathway MAD and MEDEA exhibit phenotypes similar to those of *mirror* and *Notch* loss-of-function suggesting that *dpp* may indeed be a mediator of Notch activity.

Based on these data Mirror appears to integrate the EGFR and Notch pathways during oogenesis: Gurken/EGFR signalling activates expression of Mirror in a

specific domain, which in its turn restricts expression of *fringe* in a complementary pattern. The juxtaposition of the Mirror and the Fringe domain is important for the generation of a stripe of Notch activity that results in the production and secretion of a diffusible as yet unknown morphogen. This cascade of events is essential for proper D/V axis formation and disrupting *mirror*, *fringe*, *Notch* or *dpp* activity causes D/V patterning defects.

Interplay amongst the same components has been involved in patterning of the eye (Dominguez and de Celis, 1998; Yang *et al.*, 1999). This suggests that there may be a conserved mechanism for dividing epithelia into compartments. The detailed characterisation of the nature of the above interactions and the identification of possible missing links will provide a further insight in the understanding of the mechanisms that rule organogenesis and patterning.

1.3.2.2. Mirror expression during embryonic development

mirror expression pattern in the embryo is very dynamic (McNeill *et al.*, 1997) (Figure 1.12). *mirror* loss-of-function causes embryonic lethality (McNeill *et al.*, 1997; Zhao *et al.*, 2000) indicating that its function is important for early development. The *mirror* transcript is not maternally deposited in the embryo and based on *in situ* analysis the first signs of zygotic expression are detected as early as the cellular blastoderm stage (stage 5) in an anterior ventral patch at the site of the presumptive anterior midgut invagination. At the same time expression is also detected at the site where the dorsal folds will form. Expression in these tissues persists during gastrulation and early germband extension. During germband extension and until stage 11 the anterior expression marks the site of the stomodeal invagination. The

dorsal expression persists in the folds and in the amnioserosa until stage 10. Between stage 10 and 11 *mirror* is expressed in segmentally repeated bands of ectodermal cells in a one-cell-wide pattern, adjacent and posterior to the *engrailed* expression domain. This marks the anterior domain of each segment. At the same stage *mirror* is also found in dorsal lateral cells of the ectoderm. The segmental expression lasts until the end of stage 11. Mirror is expressed in delaminating neuroblasts from about stage 10 and continues to be expressed in the ventral nerve cord and in the brain until at least stage 16. As the embryo undergoes germ-band retraction, *mirror* expression appears transiently in the proventriculus as it undergoes folding and the foregut-midgut boundary (McNeill *et al.*, 1997).

Expression of *ara* and *caup* in the embryo starts later than *mirror* (stage 11) first in the lateral epidermis and then in the proventriculus and parts of the developing brain (Gomez-Skarmeta *et al.*, 1996). This suggests that *mirror* is the only *Iroquois* member with a role in early embryonic development. Furthermore *mirror* appears to be the only *Iroquois* gene expressed in the ventral nerve cord.

Based on the dynamic nature of its expression *mirror* must be involved in various processes spanning various stages of embryonic development. Its function has not been studied in great detail but various phenotypes related to particular processes have been described. Early defects include delays in germ band retraction and dorsal closure (Helen McNeill, unpublished results). *mirror* mutant embryos also show defective head involution, are often twisted along the anterior/posterior axis and show overgrowth of the amnioserosa (Helen McNeill, unpublished results). Cuticle preps of *mirror* mutants have sparse denticles.

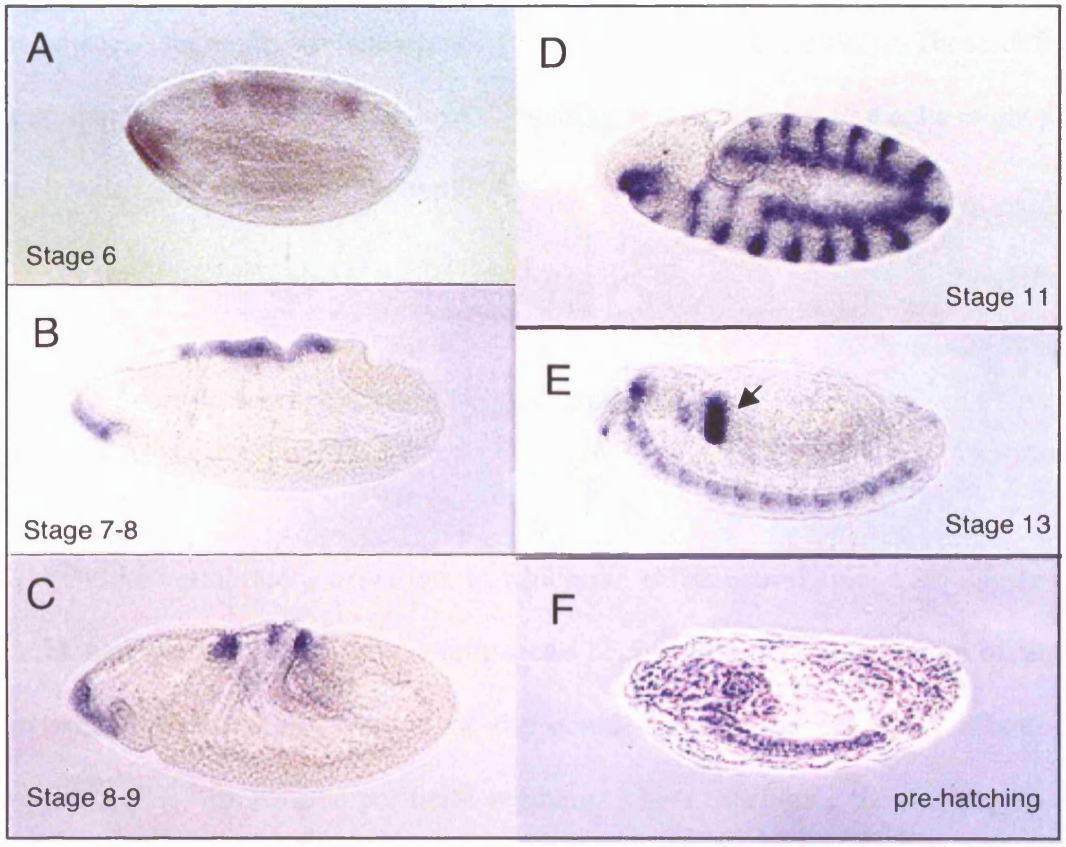


Figure 1.12: Mirror has a dynamic expression pattern in the embryo. (A) At stage 6 *mirror* is expressed in the dorsal folds and in the site of the presumptive anterior midgut invagination. During gastrulation (B) and germ band extension (C) *mirror* expression persists in the dorsal folds and the anterior midgut invagination. Expression is also detected in the amniosserosa (C). (D) From stage 11 *mirror* is expressed in a segmented pattern. (E) At stage 15 *mirror* is expressed in the developing CNS and at the proventriculus (arrow). (F) Expression persists until just before hatching. From McNeill *et al.*, 1997.

Anterior denticles are often missing from abdominal segments reflecting the lack of *mirror* expression at the anterior segmental border. In addition denticle belts from adjacent segments are sometimes fused (McNeill *et al.*, 1997). These defects suggest that the juxtaposition of *mirror* expressing and non-expressing cells might play a role in setting up the segmental border.

1.3.2.3. The role of *mirror* during CNS development

Unlike vertebrate neurulation, in which the entire neural anlage invaginates to form the neural tube, *Drosophila* neurogenesis begins with the delamination of single cells from the neurogenic region of the ectoderm into the embryo. About 30 neuroblasts (NB) are formed per hemi-segment, which represents the developmental unit of the segmented portion of the developing CNS. Each neuroblast produces a diverse population of neurons and glia, and can be identified due to its position, time of formation, and pattern of gene expression. NBs in different hemisegments that develop in the same relative position and at the same time acquire the same fate (reviewed in Skeath, 1999).

The neuroblasts in which *mirror* is expressed have been identified using a *mirror-lacZ* enhancer trap line (Broadus *et al.*, 1995; Doe, 2003). In almost all cases *mirror*-expressing NBs delaminate from the anterior region of each hemisegment, which corresponds to the segmental expression of *mirror* at stage 11.

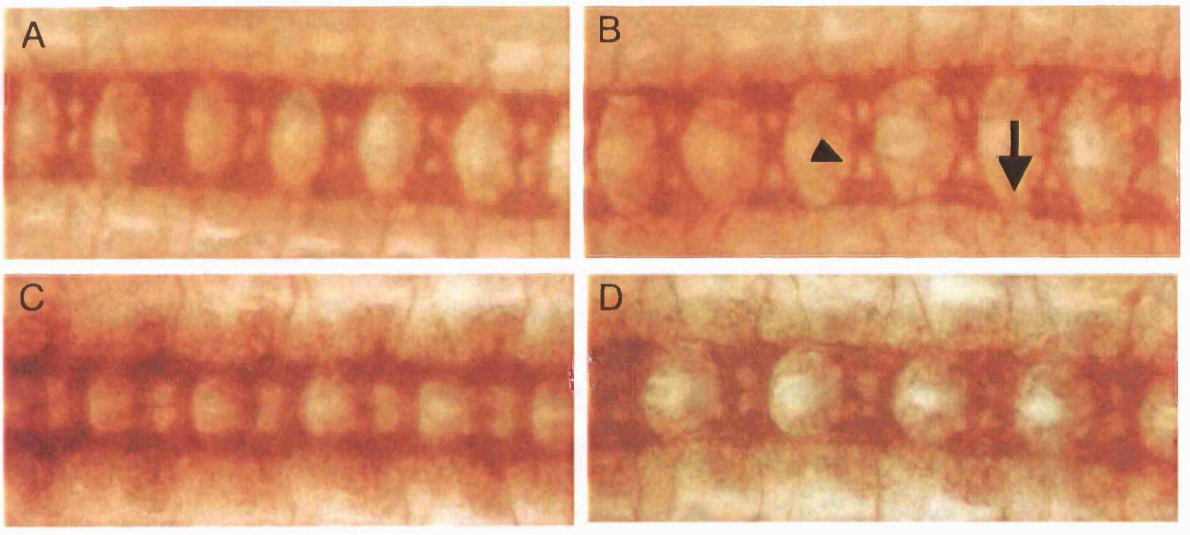


Figure 1.13: Mirror is important for CNS patterning. Embryos are stained with BP102 antibody that stains axonal tracts. (A) Wild type embryo at stage 14. Axons are organised in two longitudinal tracts connected via 2 commissures per segment. (B) In *Iroquois* mutants longitudinal tracts are disrupted (arrow) and commissures are thinner than in wild type (arrowhead). (C and D) Same views at stage 16 of development (C is wild type and D an *Iroquois* mutant). Defects seem to be less severe at later stages (compare B and D). These effects can be attributed to loss of *mirror* function since no other *Iroquois* gene is expressed in the developing CNS. All views are ventral with anterior to the right. Taken from Mohns, 2003.

The expression of *mirror* in delaminating neuroblasts, the brain and the ventral nerve cord suggests that *mirror* is important for CNS development. Neurons in the ventral nerve cord in *Drosophila* are organised in a characteristic ladder-like pattern, consisting of two longitudinal tracts connected via two commissural tracts per segment. The pattern of commissures and connectives is aberrant in *mirror* mutant embryos (Figure 1.13).

More specifically the two commissures appear to be thinner and closer together in mirror mutants at stage 12. None of these defects however is detectable at later stages in the mature, condensed CNS (M. Mohns *et al*, manuscript in preparation) suggesting that there are other genes/pathways acting to compensate for the loss of *Iroquois* during later stages. *mirror* mutants also show clear defects in the architecture of the longitudinal tracts. In wild type embryos, these are formed by three independent fascicles that run along the longitudinal axis. In *mirror* loss-of-function embryos breaks are frequently seen in the longitudinal tracts, and in some portions of the pathway only two fascicles seem to be present. In addition, regions of abnormal thickening of the longitudinal tract can be found. All these observations suggest that in the absence of *mirror* axon pathfinding is defective resulting in axons being mis-routed.

Based on a microarray experiment performed in our lab by M. Mohns, *mirror* may be causing some of these defects by directly affecting the expression of *commissureless (comm)*. Comm is a transmembrane protein, whose role in the developing CNS has been related to the guidance of the midline crossing axons (Keleman *et al.*, 2002; Seeger *et al.*, 1993). Mirror represses the expression of *comm*

in the microarray experiment and consistent with that data, expression of *mirror* and *comm* are complementary in wild type CNS. Moreover *comm* expression is expanded in *mirror* loss-of-function, suggesting that part of *mirror's* function in CNS development may be mediated by *comm* and possibly other genes implicated in axon pathfinding.

1.3.3. *mirror/Iroquois* role during wing and notum development

1.3.3.1. Outline of wing and notum development

Development of all appendages initiates during larval stages in specialised epithelial structures known as imaginal discs. The wing imaginal disc will give rise to the adult wing as well as to the dorsal mesothorax (or notum), the ventral mesothorax (or pleura) and the wing hinge region (Garcia-Bellido, 1973). The initial division of the epithelium into wing and notum territories is a result of the interplay between the EGFR and the Wingless pathways (reviewed in Klein, 2001). EGFR is activated in the notum part of the disc and *wingless* in the region of the prospective wing. EGFR activation by its ligand Vein prevents expression of *wingless* in the prospective notum. At the same time, *wingless* represses *vein* resulting in the generation of two distinct domains where negative feedback loops maintain the initial differential activation of the two signalling pathways.

Once the wing region has been assigned its specific fate a further subdivision occurs to separate the wing blade, at the centre of the disc, from the wing hinge at the periphery. The wing blade is further divided into smaller territories to delimit where

the wing veins and sensory organs will form. In the wing-proper part of the disc subdivisions between Anterior/Posterior and Dorsal/Ventral compartments result in the generation, at the compartment borders, of signalling centres important for patterning. As seen in oocyte development, a HD transcription factor, in this case *apterous,* controls expression of *fringe* to set up a narrow stripe of Notch activity that is important for subsequent patterning events. The *Iroquois* genes have not however been implicated in the repression of *fringe* expression in wing disc development.

1.3.3.2. The role of *Iroquois* in notum specification

Iroquois genes are important at various stages of wing development: At early stages they act in notum vs wing specification. They are also involved in patterning of the wing hinge and the formation of the alula, a small lobe at the posterior base of the wing. Loss of *mirror* expression from this region leads to loss of the adult structure, indicating that the Iro proteins have a direct role in the specification of the alula (Kehl *et al.*, 1998). At later stages *Iroquois* are involved in the subdivision of the notum into medial and lateral domains as well as the prepatterning of sensory organs and wing veins.

During the second larval instar, *Iroquois* are expressed in the proximal wing disc in a region that defines the prospective notum (Diez del Corral *et al.*, 1999). *Iroquois* expression in the notum is dependent on EGFR activation by Vein (Zecca and Struhl, 2002a; Zecca and Struhl, 2002b). Ectopic activation of EGFR signalling by means of expression of a constitutive active form of the receptor or an activated downstream effector (*ras*) results in ectopic expression of *Iroquois* genes (Zecca and Struhl, 2002b). Conversely clones of $EGFR^{ts}$ or ras^{-} in the notum region fail to express

Iroquois in a cell autonomous manner suggesting that there is a requirement for persistent EGFR signalling to maintain *Iroquois* expression. At the same time the distal border of the *Iroquois* domain of expression is established by repression by the Dpp signalling (Cavodeassi *et al.*, 2002) (Figure 1.14).

During third instar stage, expression of *Iroquois* genes becomes restricted to the region of the disc that will give rise to the lateral notum. *dpp* is now also expressed at the proximal most part of the prospective notum and restricts expression of the *Iroquois* in the medial notum (Cavodeassi *et al.*, 2002) (Figure 1.14). This explains why *Iroquois* clones induced during first and second instar are always associated with extensive malformations including formation of ectopic wing hinge structures, while when induced later they have no effect on the central notum and only show minor effects in the lateral regions (Diez del Corral *et al.*, 1999). These results demonstrate that there is an early requirement for *Iroquois* genes for notum specification.

Directed overexpression of any of the *Iroquois* genes in the dorsal wing pouch region using an *apterous*-GAL4 driver, results in removal of the dorsal hinge structures and malformed wings but cannot induce the formation of notum structures (Diez del Corral *et al.*, 1999). These observations indicate that Iroquois proteins on their own cannot impose a notum fate although their expression is sufficient to prevent development of the wing hinge.

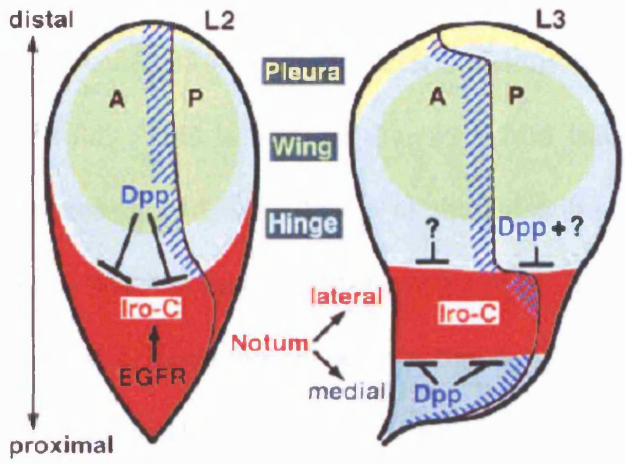


Figure 1.14: Signalling pathways acting on domain fate specification in the wing disc. During second larval instar *Iroquois* genes are activated by the Vein/EGFR pathway in the proximal region (Zecca and Struhl, 2002a and 2002b). Dpp signalling defines the notum-hinge border by restricting *Iroquois* expression in the prospective notum. During third instar stage *dpp* is also expressed at the medial notum and restricts *Iroquois* expression in the lateral notum region. Dpp and yet unidentified factors continue to maintain the notum-hinge border of *Iroquois* expression. Taken from Cavodeassi *et al*, 2002.

Expression of the *Iroquois* Complex in the notum may act to establish a border between notum and wing regions through differences in cell affinities. *Iroquois* loss-of-function clones in the eye and the wing disc have smooth borders, as if cells within the clone were trying to minimise their contacts with neighbouring cells (Diez del Corral *et al.*, 1999; Yang *et al.*, 1999). In the lateral notum *Iroquois* clones are surrounded by a visible fold. This is very similar to a fold that naturally forms between the prospective notum and wing hinge at the third instar stage. These observations in combination with the fact that *Iroquois* genes are also expressed in the dorsal folds during embryonic development and at the anterior domain of every segment imply that one of the generic function of *Iroquois* genes may be generating borders of cells based on different cell affinities. So far no downstream target has been suggested for any of the *Iroquois* members that is directly involved in cell-cell affinities but based on the above data this hypothesis is highly plausible.

1.3.3.3. *Iroquois* genes are important for bristle and wing vein formation.

The bristles are mechanosensory organs, and appear in stereotypic positions and numbers on the back of the fly. There are 11 bristles on each hemi-notum and each one of them is a cluster of four cells, all of which are progeny of a single precursor cell, called the sensory mother cell (SMC). These cells develop from proneural clusters, which are under the control of the genes of the *achaete-scute* complex of proneural genes. Expression of these genes is ruled by a network of transcription factors that generate a complex array of positional clues. The pre-pattern

of the transcription factors that make up the positional clues appears to be controlled by overlapping signalling pathways such as the Wg, Hh and Dpp pathways.

As previously mentioned, the *Iroquois* genes were named after a phenotype that causes the loss of all lateral bristles on the notum (Leyns *et al.*, 1996). *Iroquois* genes are involved in the development of 8 of the 11 macrochaetae in each hemi-notum as shown by the fact that trans-heterozygote combinations of various *Iroquois* alleles result in the loss of these bristles.

The loss of lateral macrochaetae in *Iroquois* mutants is due to the loss of the corresponding proneural clusters and SMCs in the wing imaginal disc (Kehl *et al.*, 1998; Leyns *et al.*, 1996). Ara and Caup have been suggested to directly control the expression of the genes of the *achaete-scute (ac-sc)* complex (see also chapter 5). Reduction in the *Iroquois* levels results in reduction in the levels of *ac-sc*. The effects are specific for the cells that lie within the *Iroquois* expression domain indicating that *Iroquois* genes control ac-sc expression cell-autonomously.

In addition to specifying lateral identity of the notum *Iroquois* genes also act to specify lateral identity of the bristles. If bristle formation is induced in *Iroquois* mutant background by ectopic expression of *scute,* the sensory neurons innervating these bristles adopt a medial fate –as evinced by their path towards the CNS- even though they are found in the lateral notum (Grillenzoni *et al.*, 1998). This can be viewed as additional evidence for the suggested role of the *Iroquois* genes in specifying lateral domains (Calleja *et al.*, 2002).

ara and *caup* are also important for the formation of wing veins. During wing disc development they are expressed in the regions of the wing pouch that will give rise to the veins L1, L3 and L5. Mitotic clones that lack *Iroquois* function lead to loss of vein material as well as loss of the L3 campaniform sensilla and the twin sensilla of the wing margin (TSM) proneural clusters (Gomez-Skarmeta *et al.*, 1996). The loss of the sensory organs in this area, as in the notum, is due to the loss of *ac-sc* expression. Ectopic expression of *ara* leads to the accumulation of ectopic vein material. The areas of the wing veins are delimited by the expression of *rhomboid* (*rho*). Ectopic *ara* expression also causes an expansion of *rho* expression suggesting that the effect of *Iroquois* genes on the specification of wing veins may be mediated by *rho*.

1.3.4. The role of the *Iroquois* in eye development.

1.3.4.1. Introduction in *Drosophila* eye development

The *Drosophila* eye is a compound structure that consists of approximately 800 units called ommatidia. Each ommatidium is made of 8 photoreceptor (R1-R8) and several accessory cells, namely cone and pigment cells (Wolff and Ready, 1993) . Each photoreceptor makes a single light sensitive organelle called the rhabdomere. Rhabdomeres within each ommatidium are arranged in a stereotypical pattern generating a trapezoid shape that points towards a specific direction. Based on the arrangement of the rhabdomeres within each ommatidium the whole eye field can be divided into two fields of opposing polarity. The line where these fields meet runs along the midline of the eye and is known as the equator (Reifegerste and Moses, 1999). Consistent with this terminology the dorsal and ventral margins of the eye are

referred to as the poles. Ommatidia in the dorsal half of the eye are the mirror image of those in the ventral half, i.e they exhibit opposite polarity and chirality (Figure 1.15).

Development of the eye begins during larval stages in the eye-antennal imaginal disc, which gives rise to most of the adult head structures. The eye part of the disc will form the eye and most of the head capsule while the antennal part, apart from the antenna also contributes to other structures such as the rostral membrane and the maxillary palpus (Cavodeassi *et al.*, 1999).

A complex hierarchy scheme rules specification of the eye field (reviewed in Kumar, 2001). A group of six "master control" genes (*twin of eyeless, eyeless, eyes absent, sine oculis, dachshund and eye gone*) appear to play an essential role in the process. Absence of any of these genes leads to reduction or loss of the eye field while their overexpression can cause generation of an ectopic eye (with the exception of *sine oculis*). Formation however of an ectopic eye is only possible in specific regions of a few only imaginal discs and only in the presence of an active Hh and Dpp pathway (Chen *et al.*, 1999).

The EGFR and Notch pathways have been shown to be upstream of the genes that specify the fate of both the eye and the antenna (Kumar and Moses, 2001b). Hyperactivation of several components of the EGFR pathway in the eye part of the disc causes eye-to-antenna homeotic transformations, suggesting that the EGFR pathway may act in the antenna disc to prevent the expression of genes that are essential for eye development. Expression of a dominant negative form of the Notch receptor leads to the elimination of the eye field (Kumar and Moses, 2001b; Kurata *et*

al., 2000) while hyperactivation of the Notch pathway can lead to formation of ectopic eyes in the antennal tissue (Kurata *et al.*, 2000). The effects of tampering with Notch signalling are however more difficult to interpret: In an *eyeless* mutant background increased Notch activity can result in eye-to-antenna transformations, while in wild type background the same effect can be obtained by reducing Notch activity. This suggests that, as previously proposed, Notch signalling plays a "permissive role" in specification events in which the genetic background is crucial for the outcome of its activity (Cagan and Ready, 1989).

Importantly these transformation effects can only occur within a crucial time-window during the second larval instar stage of development. This coincides with the time at which the expression of all the "master control" genes begins to overlap at the posterior portion of the eye disc (Kumar and Moses, 2001a). It had previously been reported (Postlethwait and Schneiderman, 1971) that a border of clonal restriction is formed between the two organs at earlier stages but based on recent evidence this does not mean that the two fields acquire different fates at that stage.

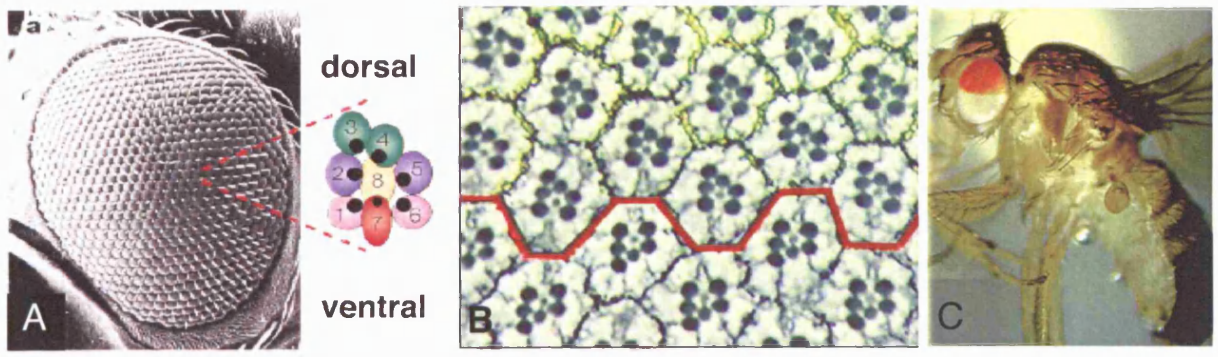


Figure 1.15: The fly eye is a compound structure. (A) Scanning electron micrograph showing that the eye consists of ~800 units, called ommatidia. Each ommatidium is made of photoreceptor and accessory cells. The photoreceptors are arranged in a trapezoid pattern as shown in the diagram. The black dot within each photoreceptor represents the light sensitive rhabdomeres. (B) Mirror image asymmetry between the dorsal and ventral halves of the eye. Ommatidia in the two fields point towards opposite directions and have opposite chirality. The line that separates the two fields is known as the equator (red line in B). (C) The Iroquois protein Mirror is expressed in the dorsal half of the eye. Enhancer trap line expressing the *mini-white* gene that results in red pigmentation, in the *mirror* pattern.

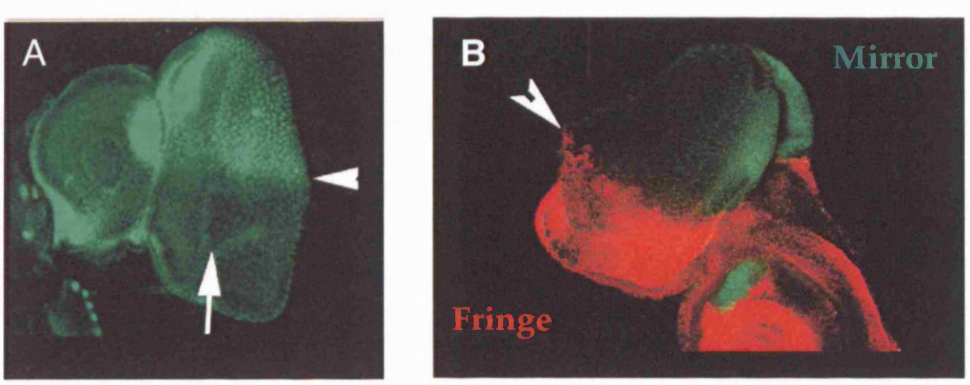


Figure 1.16: (A) *mirror* and *fringe* expression patterns in the third instar eye imaginal disc. Mirror protein (green) is expressed in the dorsal half of the disc. The morphogenetic furrow is indicated by an arrow and the midline by an arrowhead. (B) *fringe* expression is ventral as shown by anti β-Gal staining of a *fringe*-lacZ line. Taken from Yang *et al*, 1999.

1.3.4.2. Patterning of the eye and establishment of planar polarity

Differentiation of photoreceptors begins at the posterior end of the eye disc at the third larval instar stage with the formation of a physical indentation, called the morphogenetic furrow that sweeps the disc from posterior to anterior. The furrow initiates at a region of elevated Hh signalling and is dependent on the presence of Dpp and Notch signalling (Heberlein *et al.*, 1993); (reviewed in Treisman and Heberlein, 1998). Propagation of the furrow requires continuous Hh signalling. Hh is produced and secreted by newly formed photoreceptors so that undifferentiated cells ahead of the morphogenetic furrow receive the Hh signal, undergo a last round of mitosis, enter the furrow and differentiate into photoreceptors.

The first photoreceptor to acquire its specific fate is the R8 founder cell (Jarman *et al.*, 1994). R8 specification is directed by the proneural gene *atonal*, which is a target of the Hh signalling. Atonal is expressed at a narrow stripe of undifferentiated cells at the front of the morphogenetic furrow and through a process of gradual refinement its expression gets restricted to a single photoreceptor (the R8) within each ommatidium (Jarman *et al.*, 1995). It has recently been suggested that repression mediated by the homeoprotein Bar is involved in restricting atonal expression in the differentiating photoreceptors behind the furrow (Lim and Choi, 2003). Recruitment of the rest of the photoreceptors is dependent on cell-cell interactions and EGFR signalling emanating from the R8 cell via the ligand Spitz (Freeman, 1997).

Shortly after the differentiating photoreceptors emerge from the morphogenetic furrow they begin to rotate. Ommatidia in the dorsal half of the eye will rotate clockwise while those in the ventral half will rotate anti-clockwise to acquire their final orientation within the eye field. This is the readout of an as yet unidentified polarity signal commonly referred to as Factor X. According to the current model Factor X is produced and secreted at the equator and diffuses towards the poles in the form of a gradient. Its activity is interpreted by the ommatidia preclusters through binding to the Frizzled receptor generating a gradient of *frizzled* activity, which then directs their subsequent rotation. This results in two fields of opposite polarity that can easily be viewed in sections of the adult eye (reviewed in Axelrod and McNeill, 2002; Fanto and McNeill, 2004).

The establishment of planar polarity in the eye is linked to the early subdivision of the disc into dorsal and ventral territories. All three *Iroquois* genes (*mirr*, *ara* and *caup*) are expressed in the dorsal half of the disc from late first instar larval stage (Cavodeassi *et al.*, 1999; McNeill *et al.*, 1997). *fringe* is expressed during the second larval instar in the ventral half of the eye and in a pattern complementary to that of *Iroquois* expression (Figure 1.16). It has been shown that the *Iroquois* genes act to repress expression of *fringe* (Cho and Choi, 1998; Yang *et al.*, 1999). Fringe is a glycosyltransferase that modifies the affinity of the Notch receptor for its ligands Delta and Serrate resulting in a narrow stripe of Notch activity along the D/V midline, which defines the position of the equator (Papayannopoulos *et al*, 1998; Dominguez *et al*, 1998).

This localised Notch activity is then believed to lead to the induction of the polarity signal. The transmembrane protein Four-jointed is expressed at the midline

from early stages and is thought to be cleaved to generate a secreted fragment, which forms a gradient that can polarise the eye field (Zeidler *et al.*, 1999a). Absence, however, of *four-jointed* does not prevent correct establishment of polarity indicating that the activity of Factor X is also required. The atypical cadherins *fat* and *dachsous* are thought to lie upstream of factor X in the hierarchy of the polarity signalling cascade (Fanto *et al.*, 2003; Yang *et al.*, 2002) while RhoA and the JNK pathway may be downstream of Fz in interpreting the polarity signal (Fanto *et al.*, 2000; Strutt *et al.*, 1997; Weber *et al.*, 2000).

The final step in the polarity cascade is the rotation of the ommatidial clusters, which occurs in two distinct steps. The MAPK-related *nemo* acts at the first stage of rotation (0- 45°), while *roulette* (also component of the EGFR pathway) is important for completion of rotation at 90° (Choi and Benzer, 1994).

1.3.4.3. The dual role of the *Iroquois* family in patterning of the eye

The first evidence for a specific role of the *Iroquois* complex in patterning of the eye came from the observation that *mirror* loss-of-function clones in the dorsal half of the eye could induce ectopic equators (new borders of ommatidial polarity) at the interface between *mirror* expressing and non-expressing cells (McNeill *et al* 1997). Further evidence for the spatially restricted requirement for *Iroquois* gene expression came from the observation that generalised overexpression of any of the three genes led to severely diminished or no-eye phenotypes (Cavodeassi *et al.*, 1999; Cho and Choi, 1998; Dominguez and de Celis, 1998). As mentioned above the dorsal activity of the *Iroquois* genes is required for generation of the *fringe* border, which will then dictate the localised activity of Notch and establishment of the equator. Loss of *mirror*

function in the dorsal half of the eye results in ectopic expression of *fringe* while the ectopic expression of *mirror* in the ventral region severely diminishes the expression of fringe (Yang *et al.*, 1999).

It has been proposed that *mirror* and the other *Iroquois* genes also act to sharpen the equator by conferring different affinities to the cells of the dorsal compartment (Cavodeassi *et al.*, 1999; Yang *et al.*, 1999). Clones of cells that lack *mirror* or the whole *Iroquois* complex in the dorsal half of the eye have relatively smooth and round borders as if cells within the clone were trying to minimise contacts with neighbouring cells that still express the Iroquois proteins. Clones of dorsal origin that abut the D/V midline form straight borders with dorsal cells and wiggly borders with their ventral neighbours (Cavodeassi *et al.*, 1999). Furthermore trans-heterozygote combinations of *mirror* alleles that occasionally survive until adulthood display a dramatic protrusion from the surface of their eyes suggesting that cells are attempting to sort out from the neighbouring epithelium (Yang *et al.*, 1999). Such findings can be interpreted as a demonstration of a cell-sorting mechanism that contributes to the "sharpening" of the equator. One plausible model would be that *Iroquois* genes directly control the expression of some adhesion molecules that would therefore be differentially expressed between dorsal and ventral cells but these targets are yet to be identified.

In addition to establishing the equator in the eye disc, *Iroquois* genes are required for specification of the dorsal territory of the head (Cavodeassi *et al.*, 2000; Pichaud and Casares, 2000). Recently they have also been suggested to play a role in the induction of a morphologically distinct class of photoreceptors located at the dorsal margin of the eye field, known as the Dorsal Rim Area (Tomlinson, 2003; Wernet *et*

al., 2003). Misexpression of any of the three *Iroquois* genes using a GMR-Gal4 driver is sufficient to induce formation of an ectopic Dorsal Rim Area showing that *Iroquois* genes are required for the specification of this particular class of photoreceptors. This function requires Wg signalling and the effect is mediated by the HD protein Homothorax.

1.3.5. What lies upstream of *Iroquois* genes?

Expression of the *Iroquois* genes in the fly eye disc has been shown to lie downstream of the Wingless signalling pathway (Cavodeassi *et al.*, 1999). Wg is expressed at the dorsal region of the disc since the early second larval instar. Clones that lack *dishevelled* activity and therefore cannot transduce Wg signalling in the dorsal half of the eye autonomously lack *ara* and *caup* expression, while ectopic activation of the pathway using *shaggy* mutations results in ectopic *Iroquois* expression. Loss of *pannier*, which lies upstream of Wg signalling in eye disc development leads to loss of *mirror* expression while ectopic *pannier* results in ectopic *mirror* (Maurel-Zaffran and Treisman, 2000).

The Hh signalling pathway has also been suggested to lie upstream of *Iroquois* expression. Similar to *wingless*, *hedgehog* is also expressed at the dorsal region of the eye disc from late first/early second larval instar. Clones of cells in which the Hh pathway has been impaired lose expression of *Iroquois* genes while constitutive activation of the pathway results in ectopic *Iroquois* expression (Cavodeassi *et al.*, 1999).

Finally the JAK/STAT pathway has been implicated in regulation of *Iroquois* expression in the developing eye. Unpaired, the ligand required for activation of the pathway is expressed at the posterior margin of the disc adjacent to the optic stalk and is necessary for correct positioning of the equator. When *unpaired* function is lost in clones, the eye field becomes dorsalised and the equator is shifted ventrally. This shift is marked by expansion in expression of *mirror* indicating that the JAK/STAT pathway is another candidate for control of *Iroquois* expression and correct establishment of the equator (Zeidler *et al.*, 1999b).

Currently there is no information as to what inhibits *Iroquois* expression in the ventral region. It is believed that the Polycomb (Pc-G) proteins are involved in the maintenance of the repressed state for many homeotic genes. A reduction in the dosage of the Pc-*G* gene products can cause a relief of the ventral repression in *Iroquois* enhancer trap lines suggesting that Polycomb members may be involved in the maintenance of this ventral repression (Netter *et al.*, 1998).

1.4. Aim of the project- Outline of the thesis

The aim of my project was to investigate the function of the Iroquois protein Mirror at the level of transcription regulation. To do so I first undertook an *in vitro* assay to characterise Mirror DNA binding specificities (described in chapter 2). Then I went on to confirm the results of the site selection by showing that binding of Mirror to this novel DNA site is specific and requires a short consensus motif that is part of a wider sequence identified by means of the *in vitro* assay (discussed in chapter 3).

I compared the binding specificities of other members of the *Iroquois* family to that of *mirror* and showed that specificities seem to be conserved amongst *Drosophila* and vertebrate homologues. I also demonstrated that these specificities are distinct from that which was previously suggested on the basis of a DNase I protection assay performed for one of the *Drosophila* members (discussed in chapter 4).

I used the *in vitro* identified site in functional assays to show that in the context of *Drosophila* development it can mediate transcriptional control. More specifically, in the context of eye development, binding of Mirror (or the other Iroquois) to this site causes transcriptional repression in an *in vivo* reporter assay (described in chapter 5). Using this information together with the results of a microarray analysis designed to identify direct *mirror* targets we have investigated ways of combining data from the *in vitro* studies, the genome wide approach and *in silico* analysis to obtain improved lists of potential mirror downstream targets. The advantages and drawbacks of these approaches and the case for a candidate Mirror target are presented in chapter 6.

Chapter 2:

DNA binding site selection assay for the *Drosophila* Iroquois protein Mirror

As discussed above, HD transcription factors have been very extensively studied with regard to their DNA binding affinities. Iroquois (Iro) proteins belong to the wider group of atypical HD transcription factors of the TALE class. Several lines of evidence suggest that Iroquois proteins should have different DNA binding specificities than classic HD transcription factors.

- First, unlike classic HOX proteins, which have a 60 amino acid HD, TALE family members have 63 amino acid in their HD with a **T**hree **A**mino acid **L**oop **E**xtension between helix 1 and helix 2 (Burglin, 1997).
- Second, the residue at position 50, which has been shown to be critical for binding specificities, (Hanes and Brent, 1989) is of a different nature within the TALE class. In classic Homeodomains a polar residue such as Glutamine (Q) occupies position 50, establishing direct contact with the DNA whereas in the TALE class a small non-polar residue is found at this position. In the case of the Iroquois proteins this residue is an Alanine (A).
- Third, PBX and MEIS transcription factors that also belong to the TALE class have distinct binding sites, neither of which coincides with the classic HOX motif. All the above suggests that the DNA

preferences of members of the TALE class might be of a different nature.

In our effort to understand the role of Mirror, one of the three Iroquois family proteins in *Drosophila*, and its function at the level of transcriptional control of downstream targets, we set out to characterize its DNA binding specificities. To do so we decided to undertake an unbiased approach for identifying the DNA binding preferences of Mirror and potentially those of the whole Iroquois complex.

2.1. DNA site selection assay: The principle

To identify the Mirror DNA binding sequence we decided to perform an *in vitro* assay based on the intrinsic affinity of transcription factors for DNA. This assay has successfully been applied in the past to characterize binding specificities of other transcription factors (Pollock and Treisman, 1990). It relies on the enrichment of a pool of ^{32}P-labelled oligonucleotides for sequences that are specifically bound by the protein of interest. These sequences are then selected through rounds of immunoprecipitation (IP). The oligonucleotides used for this experiment consist of a random core of 26nt flanked by two stretches of known sequence (25nt) to enable PCR amplification of the selected sequences. After mixing the pool of oligonucleotides with the protein, immunoprecipitation is used to isolate specific protein-DNA complexes. The DNA is then amplified by PCR and subjected to further rounds of selection. After sequential rounds of selection, complexes are subjected to an electrophoretic mobility shift assay (EMSA). Specific bands start appearing on the autoradiography film after the first couple of selection cycles. These become stronger after subsequent rounds and can be cut out of the gel. The selected DNA is then

recovered, amplified by PCR and finally cloned and sequenced. The outline of the selection assay is shown in Figure 2.1.

The exact number of rounds after which one can expect a significant enrichment of the initially random pool of oligos can only be determined experimentally as it depends on the affinity of the protein for its binding site and the efficiency of the immunoprecipitation. Both too few or too many cycles can affect the outcome of the experiment as in the first case there won't be enough specificity and in the latter there might be an artificial bias for very high affinity sites. In practice one has to arbitrarily choose a number between 3-5 rounds, sequence the selected sites, look for consensus motifs and assess the results. If there is no obvious consensus motif it may be necessary to modify the conditions of the binding reaction and/or the immunoprecipitation step, or increase the number of cycles.

The Mirror protein that was used for this assay was produced in a cell free transcription-translation system. Our attempts to produce full length Mirror protein in bacteria had not been successful and the only construct that could be made at a satisfactory yield was a partial, C-terminal construct lacking the Homeodomain. The *in vitro* translated construct was on the other hand produced at a satisfactory yield and could be recognised by Mirror-specific antibodies (Figure 2.2 and 2.5).

To ensure that we would be truly selecting sequences specific for Mirror and to reduce the possibility of false positives I performed the site selection twice using different antibodies for the immunoprecipitation step. The first time I used an α-Mirror affinity purified peptide antibody and protein-A sepharose beads for the immunoprecipitation.

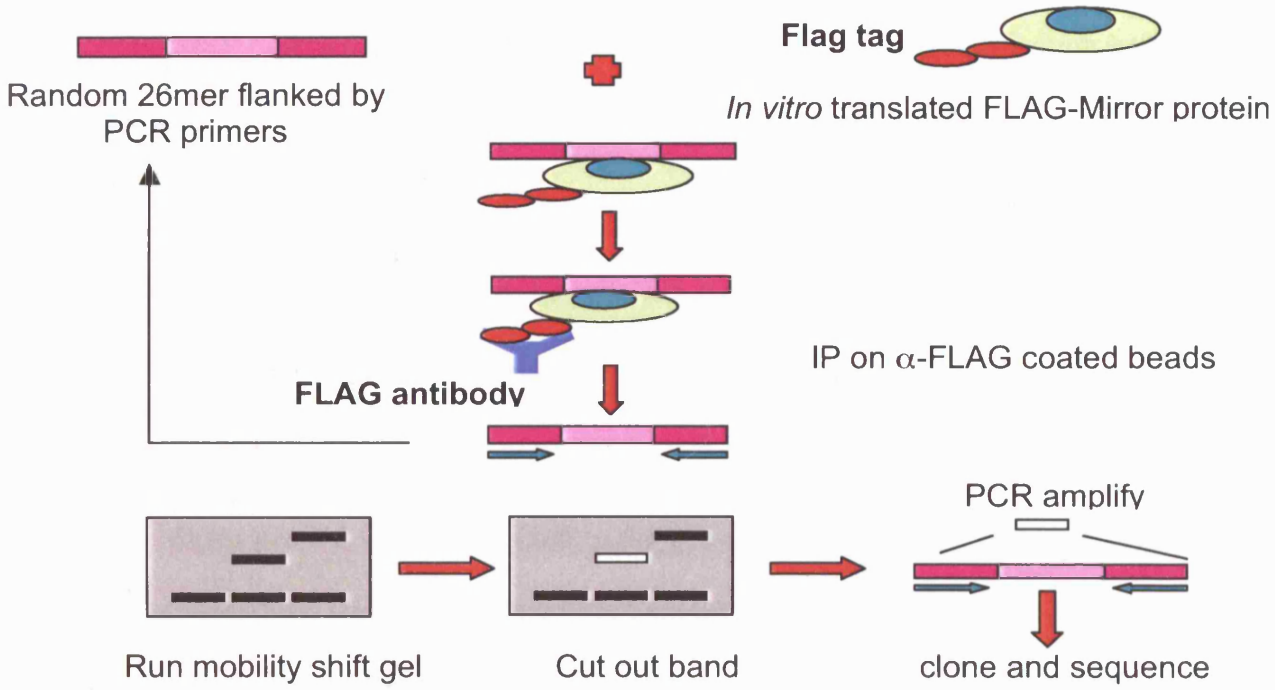


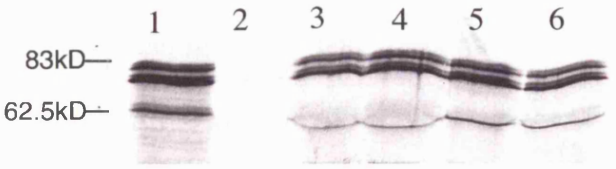
Figure 2.1: Schematic representation of the site selection assay (here shown with the αFLAG antibody): *In vitro* translated FLAG-Mirror protein was incubated with a pool of oligonucleotides. consisting of a random 26mer flanked by two stretches of known sequence. The oligos were labelled with ^{32}P to allow monitoring of the process. Protein–DNA complexes were purified by IP on αFLAG-coated beads. Selected DNA was eluted, amplified by PCR and used for a total of 4 rounds of selection. Complexes were then subjected to an EMSA. DNA was recovered from the gel, amplified by PCR and finally cloned and sequenced.

I then subcloned full length Mirror into a FLAG-vector (pFTX9, gift of C.S Hill) and repeated the experiment using α-FLAG conjugated agarose beads for the IP. In both cases I carried out 4 full cycles of binding, IP and PCR amplification before performing the EMSA. The results of the two independent experiments will be discussed in the following sections.

2.2. Site selection I : α-Mirror Immunoprecipitation

Mirror protein was synthesized in the *in vitro* transcription-translation system (IVT) and synthesis was verified by western analysis. The α-Mirror antibody used for the site selection was then tested for its ability to specifically immunoprecipitate *in vitro* translated Mirror protein (Figure 2.2). I carried out 4 rounds of selection and used samples from all rounds on the electrophoretic mobility shift gel. Specific bands started appearing after round 2 and as expected they became stronger after subsequent rounds (Figure 2.3).

Interestingly we could see two different bands in each reaction. Addition of the antibody into the reaction also resulted in a "double" supershift implying that both bands were specific. We speculate that the reason for the appearance of these 2 bands was that some of the selected sequences probably contained either a second complete binding site, or some lower affinity sites. In these cases two protein molecules would be binding to the same oligonucleotide, causing a further retardation in the mobility of the DNA.



1:*In vitro* translation reaction
2: IP with αMAX ab (neg control)
3: IP with αMirr Cterm ab
4:IP with αMirr Cterm IgG
5: IP with αMirr Nterm ab
6: IP with αMirr Nterm ab (aff. purified)

Figure 2.2: IP with various α-Mirror antibodies (performed by Trevor Littlewood). The antibody used for the site selection experiment was an α-Mirror N-term affinity purified peptide antibody (lane 6). In the EMSA the α-Mirror C-term antibody (lane 3) was used as an additional control.

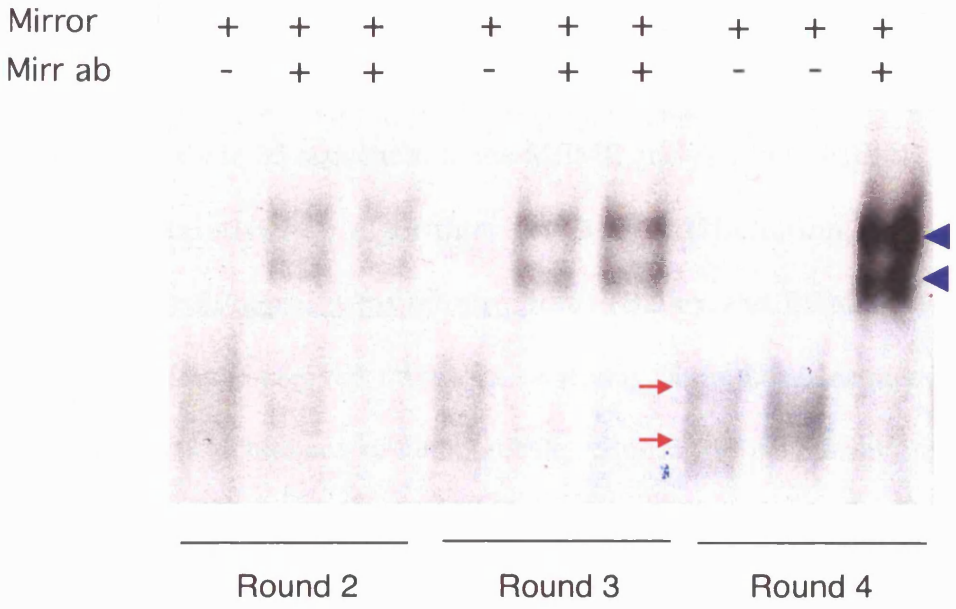


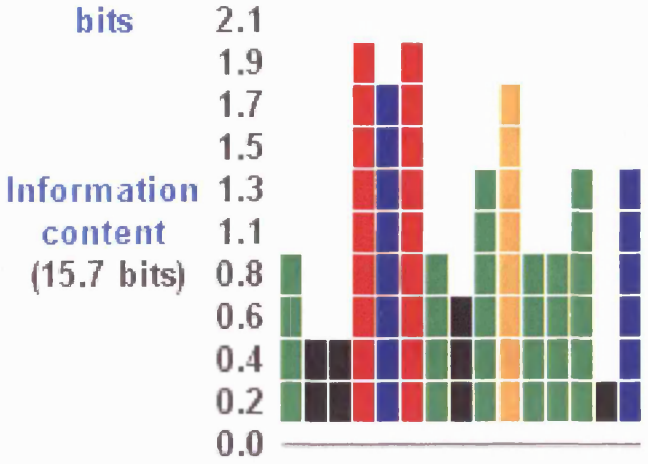
Figure 2.3: Site selection using the αMirror antibody for the IP. EMSA with samples from the last 3 rounds of the site selection experiment. The arrows indicate protein-DNA shifts and the arrowheads antibody supershifts. The enrichment is more evident in the case of the supershifts (compare rounds 2, 3 and 4). For rounds 2 and 3 two different αMirror antibodies were used for the supershifts (see figure 2.2). For round 4 the Mirror-DNA sample was loaded twice, due to a leak. Round 4 bands were excised from the gel to isolate the selected DNA (see text).

The outcome of this would be two populations of protein-DNA complexes: one with a single protein molecule and a second with two protein molecules bound to the DNA. Addition of the antibody generates a higher order complex that is expected to have a stabilizing effect, which explains why bands seem to be more robust in the presence of the antibody.

I isolated DNA from both bands after the 4th round of selection, amplified it, cloned it into a PCR-cloning vector (TOPO-TA kit, Invitrogen) and sequenced 40 individual colonies using sequencing primers specific for the vector. I obtained good quality sequencing from 35 of these colonies and used these for subsequent analysis.

I submitted these 35 sequences to the MEME analysis tool (MEME: **M**ultiple **E**xpectation /Maximisation algorithm for **M** otif **E**licitation; available at http://meme.sdsc.edu/meme/website/intro.html) (Bailey and Elkan, 1994). This programme identifies conserved motifs in relatively short DNA sequences using statistical modelling techniques to automatically choose the best width, number of occurrences, and distribution of motifs in the input sequences. Because we did not want to bias the analysis we initiated the search using the default settings i.e allowing for identification of motifs that were equal to or longer than 6 nucleotides with no restriction in the number of occurrences or their distribution within the 26nt core.

A



B

Multilevel consensus sequence

TTAACATATGTTTTC
AGG CGA AAAA
AT G

C

NAME	STRAND	START	P-VALUE	SITES
419	+	4	3.25e-07	AGA TAAACATATGTTTAC GAATTATC
405	+	7	4.27e-07	ATTTAT AGGACACGTGTTTTC CTTCC
107	+	3	6.04e-07	TA TATACACATGTTTTC GGTTTGTC
102	-	5	9.64e-07	ACCAGAG ATAACATATGTTAGC CAGA
406	-	7	2.03e-06	CTACA TTAACATGAGAATTC ATGTTA
117	+	12	3.52e-06	GAGCAACAAC ATGACAAGTGTATGC
411	-	5	6.86e-06	GCGATCA TTTACATATGTAATA AGAT
401	-	9	7.94e-06	CGC AGAACATTTGATTAC TCACTCTC
417	+	12	3.19e-05	CAACTGGGTC TGGACATATGCATCG
114	-	1	3.19e-05	ACACATCTCC TTTATATGACTTTTC

Figure 2.4 : MEME analysis of site selection 1. The sequences obtained from the selection assay with the αMirror antibody were submitted to the MEME analysis tool for identification of a consensus motif. (A) Information Content Diagram showing the degree of conservation at each position of the motif. The height and colour of each column corresponds to the frequency of the prevalent base at each position of the motif. Each of the four bases has been assigned a different colour (A= red, C=blue, G=yellow, T= green) If no base has frequency above 0.5, the column in the diagram is coloured black. (B) The multilevel consensus motif is an aid in understanding and remembering the motif. For each column the bases are arranged in decreasing order of probability with the most probable base on top. (C) List of sequences aligned to highlight the emerging motif.

MEME motifs are presented as position–specific probability matrices showing the probability of each one of the four bases at each position of the motif. MEME automatically rates motifs using e-value as an index of their statistical significance. The e-value of a sequence is the expected number of sequences in a random database of the same size that would match the motif as well as this given sequence does. Motifs with the lowest e-value are identified first.

MEME also provides an information content diagram of each motif that indicates which positions in the motif are more highly conserved (Figure 2.4). Each vertical column within this diagram is characterised by the amount of information it contains relative to the background noise. Highly conserved positions contain high level of information, whereas positions where there is no nucleotide preference have low information. This information is depicted by the height and the colour of each column: Each of the four bases has been assigned a separate colour. When the frequency of occurrence is above a statistically significant threshold the colour of the column corresponds to the colour of the most likely base for this position of the motif. Positions at which no base reaches this threshold are coloured black (Figure 2.4).

Summing the information content of each position gives the total information content (IC) for a given motif (measured in bits and shown next to the information content diagram, Figure 2.4). The Information Content is a measure of the usefulness of the motif for database searches. In other words it defines if it can be considered as a consensus motif. For a motif to be useful for searching databases of sequences it must contain at least $\log_2(N)$ bits of information , where N=the number of sequences in the database being searched. For example, to effectively search a database containing

100,000 sequences for occurrences of a *single* motif, the motif should have an IC of at least $\log_2(100{,}000)=16.6$ bits. For this site selection experiment $\log_2(35)= 5.129$ while the IC was calculated to be 15 bits, indicating the sequence identified by means of the site selection represents a consensus motif.

The results of the MEME search are most easily viewed in the form of the multilevel consensus sequence, which shows nucleotides at each position of the motif in order of decreasing probability. The list of sequences submitted to the MEME tool, the multilevel consensus sequence and the information content diagram are shown in Figure 2.4.

2.3. Site selection 2: α-FLAG Immunoprecipitation

For the second site selection experiment I used *in vitro* translated FLAG-tagged Mirror protein and FLAG-coated agarose beads for the immunoprecipitation step. Prior to initiating the experiment I tested the α-FLAG antibody for its ability to immunoprecipitate Mirror protein (Figure 2.5).

I used the same conditions for the binding reaction and immunoprecipitation as for the first experiment and carried out 4 rounds of selection. Running the samples on the mobility shift gel I could detect faint bands appearing after the second round and isolated the protein-DNA complexes after round four. Similar to the first site selection two bands were forming for each sample both in the case of the protein-DNA shift as for the antibody supershift (Figure 2.6).

I picked 50 random colonies and obtained good quality sequencing for 46. I submitted these sequences to the MEME analysis tool using the same search parameters as previously. The obtained consensus motif is shown in Figure 2.7. Based on the number of sequences that contributed to the consensus motif the second site selection appears to have been more successful in identifying the Mirror binding motif. The Total Information Content was in this case equal to 17, with $\log_2(46)=5.52$ indicating that the results of the second experiment had a higher statistical significance.

Figure 2.5: IP with αFLAG-coated beads immunoblotted with the αFLAG antibody. *In vitro* translated Mirror can be immunoprecipitated using the αFLAG antibody (arrow). After addition of the *in vitro* translation reaction and subsequent washes the beads were boiled in SDS buffer and spun down. Both the supernatant (eluate) and pellet (beads) were loaded on the gel. The two asterisks indicate the two IgG bands whose presence is due to the fact that the same antibody was used for IP and Western blotting.

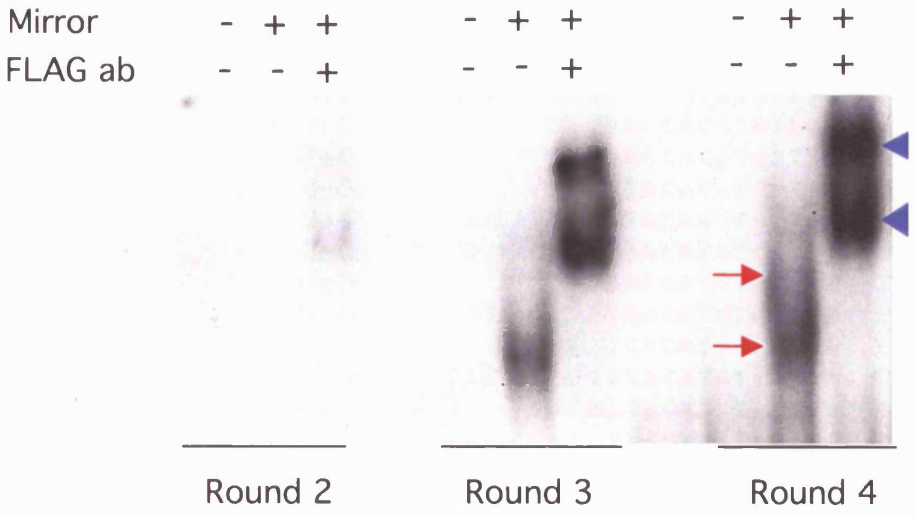
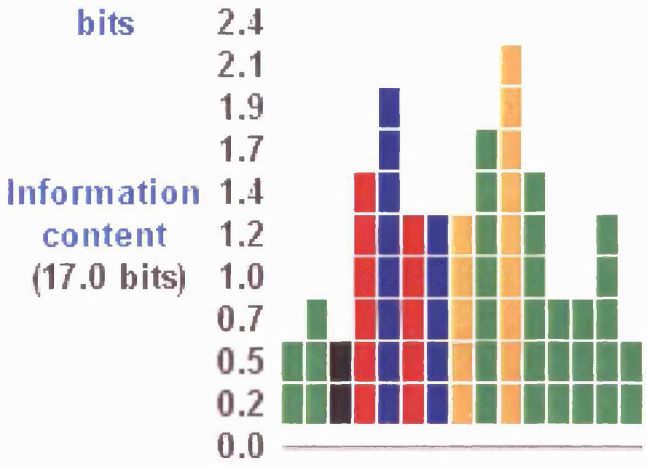


Figure 2.6: Site selection using the αFLAG antibody for the IP. EMSA with samples from the second, third and fourth round of site selection. Faint bands start appearing after round 2 and become stronger after subsequent rounds. The arrows indicate the Mirror-DNA shifts and the arrowheads the antibody supershifted complexes. Bands from round 4 were excised from the gel to isolate the DNA.

A



B

Multilevel consensus sequence

TTAACACGTGTTTTT
A T TA AA

C

NAME	STRAND	START	P-VALUE	SITES
53	-	2	3.60e-08	AGTGAAT ATAACACGTGTTATT T
23	+	8	3.60e-08	AGTGAAT ATAACACGTGTTATT T
55	-	6	5.66e-08	GC TTTACACATGTTTTT GTGTT
48	+	9	5.66e-08	TGATCATA TTTACACGTGTAATT GTC
43	-	7	5.66e-08	AGA TTAACACGTGTAATT GTGACA
41	+	3	5.66e-08	AC TTTACACATGTTTTT GTGTT
40	+	9	5.66e-08	TGATCATA TTTACACGTGTAATT GTC
35	-	6	5.66e-08	GC TTTACACATGTTTTT GTGTT
25	-	6	5.66e-08	AC TTTACACATGTTTTT GTGTT
22	-	6	5.66e-08	GC TTTACACATGTTTTT GTGTT
18	-	6	3.80e-07	TTAACACGTGTATTA CTTGT
42	+	12	4.53e-07	ATGAACATGC TTAACATATGTTTTT
19	+	7	5.03e-07	TAGTTG TTAACACGTGTAATC AGAGT
34	-	3	7.57e-07	GCCGTTTCC AGAACACGTGTTTTA CA
45	+	3	1.10e-06	GG TTTACACGTGTTTAC ACTGCGATT
49	+	9	1.22e-06	TAACATGT GGTACACGTGTTTTA TA
47	+	9	1.22e-06	NGGAAGAC ANAACATGTGTTTTT ACA
50	+	6	1.91e-06	ATNTG AAAACATGTGTTTTT CTCCA
14	-	4	1.91e-06	TGGAAGAC AAAACATGTGTTTTT ACA
27	+	4	4.20e-06	TCG TTAACATATGTTAAT TTGACAAG
57	-	4	4.52e-06	CACTGTAA TTTACATGTGTTTAG TTA
21	+	4	6.53e-06	TGT AGTACAAGTGTATTT AGATCACT
30	+	8	8.06e-05	ATACAAG GTTAGTCATGTAATT TAAC
31	-	4	1.50e-04	TTTATATC ATAACGCGTGAGATG CCT
10	+	4	7.54e-04	TTG GGGNNACGTATTATT GTAAGAATG

Figure 2.7: MEME analysis of site selection 2. (A) Information content diagram and (B) multilevel consensus sequence showing that the consensus motif identified from sequencing the selected sequences from the αFLAG IP is very similar to the one emerging from the α-Mirror IP (compare to figure 2.6). (C) List of sequences selected by means of the αFLAG IP. For more information on the information content diagram and multilevel consensus refer to figure 2.4 and the text.

2.4. Comparing the two site selection experiments

The consensus motifs obtained from the two experiments were very similar. Overall it appears that the α-FLAG IP has been more successful in selecting Mirror specific sequences as shown by the fact that more of the input sequences contained the suggested Mirror binding site. In both cases we could find sequences that did not contain the motif. This shows that the immunoprecipitation step is not 100% efficient and that depending on the quality of the antibody some non-specific sequences may be expected to remain in the pool even after 4 rounds of selection.

As an additional source of information I carried out a MEME search using sequences obtained from both experiments (a total of 81) and the results are shown in Figure 2.8. The IC of the combined search was equal to 17.7, indicating that the combined search was more successful in identifying a true consensus motif.

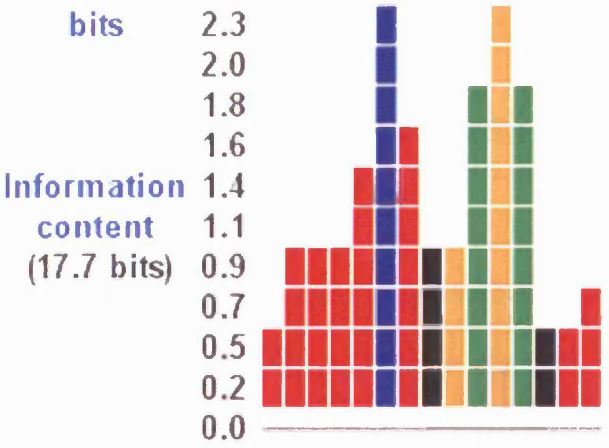
Another way of viewing the results of the site selection and using them to identify a consensus motif is a positional weight matrix. Transcription factor binding sites are sometimes variable in their sequences and allow some degeneracy. In addition to this, the IP step in the site selection assay is not very efficient and some non-specific sequences are still present after 4 selection rounds. Positional Weight Matrixes provide information as to the frequency of each base at each position of the alignment. The output of the results can be viewed as an alignment matrix showing the number of occurrences of nucleotide *i* in position *j* of the motif. The alignment matrix for site selections 1 and 2 is shown in table 1.

A	**28**	**38**	**39**	**43**	**49**	1	**45**	3	19	8	10	11	24	**33**	**39**
C	11	9	0	4	6	**62**	4	**30**	0	17	23	6	13	26	11
G	**28**	6	0	9	11	14	3	19	**43**	7	**44**	13	14	10	9
T	10	24	38	21	11	0	25	25	15	**45**	0	**47**	**26**	8	18
	A/G	**A**	**A**	**A**	**A**	**C**	**A**	**C**	**G**	**T**	**G**	**T**	**T**	**A**	**A**

Table 1: Position Weight Matrix for identification of a consensus Mirror motif from site selections 1 and 2. Each element in the matrix represents how many times nucleotide *i* was found in position *j* of the alignment (Sosinsky *et al.*, 2003); http://trantor.bioc.columbia.edu/Target_Explorer). Frequencies of prevalent bases at each position are shown in bold. At the bottom of the table is shown the emerged consensus motif.

From both the MEME and the Position Weight Matrix analyses it becomes evident that despite some minor flaws, the *in vitro* site selection has indeed identified a Mirror consensus motif [A/G]AAAACACGTGTTAA.

A



B

Multilevel consensus sequence

```
AAAAACACGTGTTAA
G  TT    TA    A T
```

C

NAME	STRAND	START	P-VALUE	SITES
55	+	6	2.45e-08	AACAC **AAAAACATGTGTAAA** GC
41	-	3	2.45e-08	AACAC **AAAAACATGTGTAAA** GT
35	+	6	2.45e-08	AACAC **AAAAACATGTGTAAA** GC
25	+	6	2.45e-08	AACAC **AAAAACATGTGTAAA** GT
22	+	6	2.45e-08	AACAC **AAAAACATGTGTAAA** GC
53	+	2	6.09e-08	A **AATAACACGTGTTAT** ATTCACT
23	-	8	6.09e-08	A **AATAACACGTGTTAT** ATTCACT
43	+	7	8.12e-08	TGTCAC **AATTACACGTGTTAA** TCT
42	-	12	1.20e-07	**AAAAACATATGTTAA** GCATGTTCAT
48	-	9	1.39e-07	GAC **AATTACACGTGTAAA** TATGATCA
40	-	9	1.39e-07	GAC **AATTACACGTGTAAA** TATGATCA
19	-	7	1.60e-07	ACTCT **GATTACACGTGTTAA** CAACTA
18	+	6	2.88e-07	ACAAG **TAATACACGTGTTAA**
45	-	3	3.37e-07	AATCGCAGT **GTAAACACGTGTAAA** CC
107	-	3	4.81e-07	GACAAACC **GAAAACATGTGTATA** TA
50	+	5	6.02e-07	ATNT **GAAAACATGTGTTTT** TCTCCA
34	+	3	7.50e-07	TG **TAAAACACGTGTTCT** GGAAACGGC
14	+	4	7.50e-07	TGT **AAAAACACATGTTTT** GTCTTCCA
47	-	9	8.23e-07	TGT **AAAAACACATGTTXT** GTCTTCCX
405	-	7	1.47e-06	GGAAG **GAAAACACGTGTCCT** ATAAAT
27	-	4	1.74e-06	CTTGTCAA **ATTAACATATGTTAA** CGA
419	+	3	3.04e-06	AG **ATAAACATATGTTTA** CGAATTATC
411	+	5	3.49e-06	ATCT **TATTACATATGTAAA** TGATCGC
57	-	5	3.75e-06	CACTGTA **ATTTACATGTGTTTA** GTTA
21	-	4	4.58e-06	AGTGATCT **AAATACACTTGTACT** ACA
102	+	5	5.22e-06	TCTG **GCTAACATATGTTAT** CTCTGGT
49	-	9	5.56e-06	TA **TAAAACACGTGTACC** ACATGTTA
117	-	12	2.17e-05	**GCATACACTTGTCAT** GTTGTTGCTC
406	+	7	2.34e-05	TAACAT **GAATTCTCATGTTAA** TGTAG
401	+	9	4.46e-05	GAGAGTGA **GTAATCAAATGTTCT** GCG

Figure 2.8 : MEME analysis of the results from site selection 1 and 2 Sequences obtained from the two experiments were submitted to the MEME tool. Information content diagram, multilevel consensus and list of sequences are shown. For more information on interpreting the results refer to figure 2.4 and the text.

2.5. Testing the newly identified binding site in EMSAs

To experimentally confirm these results I used oligonucleotides that contained the MEME motif obtained from both experiments and tested if Mirror protein could indeed bind this sequence (Figure 2.9). Addition of *in vitro* translated Mirror to a labelled oligonucleotide carrying the AAAAACACGTGTTAA motif results in a shift in the mobility of the free oligo. Binding of Mirror to this sequence is specific, as verified from addition of an anti-FLAG antibody that specifically supershifts the Mirror-DNA band. The presence of Mirror protein in the reaction was confirmed by Western blot using both an anti-Mirror and an anti-FLAG antibody. The negative control, an *in vitro* translation reaction that contained no DNA template for Mirror did not give a band of that size.

Note that there is a non-specific band common to all samples that is not affected by the presence of Mirror protein, nor addition of the antibody. This band was subsequently seen in all protein DNA binding reactions, irrespective of the identity of the protein or the sequence of the oligo. We therefore believe it corresponds to a protein in the reticulocyte lysate that binds non-specifically to DNA and it will be indicated in all figures by an asterisk.

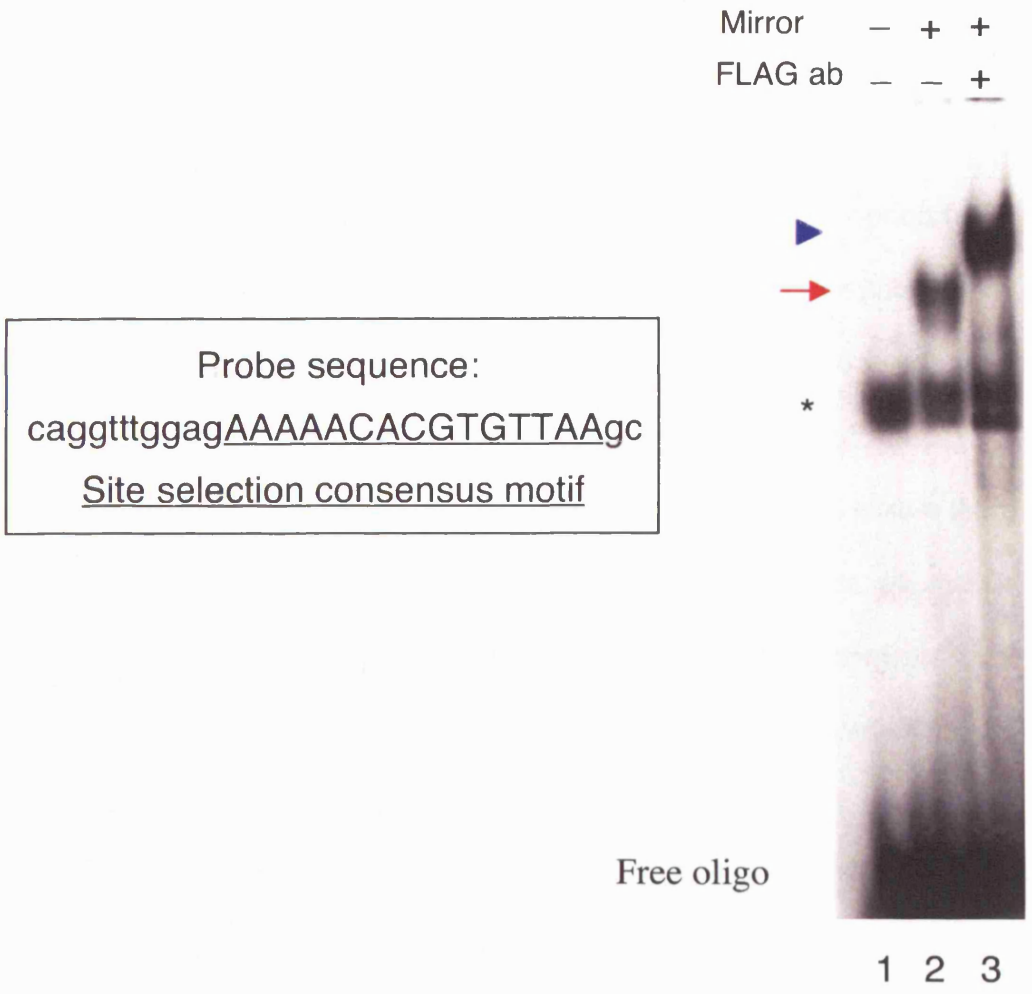


Figure 2.9: Mirror can bind the site selection consensus motif. Full length Mirror protein was mixed with a labelled probe that carried the consensus motif that emerged from the combined results of the two site selection experiments (see box for sequence of the probe, underlined is the site selection consensus motif); From this experiment only it is not possible to conclude that binding is due to specific recognition of a consensus motif. Validation of the site selection results will be presented in chapters 3 and 4. Lane 1, mock translation reaction (no Mirror protein); lane 2, *in vitro* translated Mirror; lane 3, *in vitro* translated Mirror mixed with αFLAG antibody. The arrow indicates the protein-DNA shift and the arrowhead the antibody supershift. The asterisk indicates a non-specific band.

2.6. Concluding remarks:

The first step in understanding the role of Mirror as a transcription factor was to identify its DNA binding specificities. In this chapter I present an unbiased *in vitro* approach to identify the Mirror binding site. I have conducted a site selection experiment in which immunoprecipitation is used to enrich a pool of initially random oligonucleotides for sequences that bind specifically to Mirror. To ensure that I was selecting *bona fide* specific sites I performed the experiment twice, using different antibodies for the Immunoprecipitation. The results of the two experiments were very similar and the occurrence of a consensus motif was evident by eye and confirmed by use of two different consensus identification tools:

- The MEME analysis tool (http://meme.sdsc.edu/meme/website/intro.html)
- The Position Weight Matrix at the Target Explorer site (http://trantor.bioc.columbia.edu/Target_Explorer)

Both programmes identified the same consensus motif [A/G]AAAACACGTGTTAA indicating that our assay has been successful in selecting a Mirror binding site. I have confirmed by EMSAs that Mirror can specifically bind this sequence. In the following chapters I will present experiments aiming at further characterising the newly identified site, both *in vitro* and *in vivo*.

Chapter 3:

Characterisation of the Mirror binding site

In the previous chapter I described an assay aimed at identifying DNA sequences based on their affinity for Mirror protein and showed that Mirror can bind the motif identified by means of this selection protocol in EMSAs. In the following sections I present a series of experiments aiming at identifying the minimum requirements for Mirror binding to DNA. Several features of the consensus motif were tested either by comparing the efficiency of the binding in direct competition assays or by comparing the intensity of bands obtained in EMSAs where equivalent amounts of labelled probe and protein were used. In all cases where a competition assay was performed this will be clearly stated to distinguish them from the more qualitative observations based on comparison of band intensity. No dissociation constants were calculated for the protein-DNA complexes and the words " affinity and efficiency of binding" are used interchangeably, mentioning however the experimental means of acquiring the data.

3.1. What elements of the site selection motif are essential for binding

Submitting the sequences obtained from both site selection experiments to the MEME analysis tool and using a Position Weight Matrix we obtained the same consensus: **A(g)-A-A(t)-A(t)-A-C-A-C(t)-G(a)-T-G-T-T(a)-A-A(t).** Letters in lower case indicate nucleotides that appeared in this position of the motif with lower

frequencies. Closer inspection of the motif revealed that it maintained the main characteristics of the motifs identified through the two independent experiments:

- a short motif (ACA) and its inverted repeat (TGT) that had the highest scores in both the MEME and the PWM consensus.
- 2 nucleotides that were not strictly specified between the ACA and TGT, with a small preference for a CG pair
- an A/T rich region on both sides of the ACA and TGT motifs

Based on these observations I decided to test the requirement for these 3 features of the motif in EMSAs and assess their importance for Mirror binding.

3.1.1. The AT rich ends are dispensable for binding

I first tested whether the A/T rich flanking region was essential by placing the ACACGTGT motif in different contexts. I found that Mirror can bind the ACACGTGT sequences in the context of a number of different flanking sequences in EMSAs (Figure 3.1). The complexes can be supershifted by the addition of antibodies to Mirror or to the FLAG tag demonstrating that an ACACGTGT sequence is sufficient for specific Mirror binding even in the absence of the AT-rich neighbouring regions. Although this is not evident in Figure 3.1, in direct competition assays binding seems to be slightly more efficient in an AT rich context (data not shown).

3.1.2. The nucleotides separating the ACA and TGT motifs act as a spacer

I then went on to test whether changing the nucleotide composition of the two central positions of the motif had an effect on binding. I changed the central two nucleotides from CG to TA (i.e I varied the motif from ACACGTGT to ACATATGT keeping the same flanking sequences) and tested Mirror binding in EMSAs. Mirror binds both the ACACGTGT and the ACATATGT motifs and binding is specific as shown by means of the antibody supershift (Figure 3.2A).

In both cases the nucleotides separating the ACA and TGT palindromic motifs are contributing to the formation of an "extended" palindrome, i.e ACAC-GTGT or ACAT-ATGT. To test if this was a requirement for binding I tested Mirror binding to ACAnnTGT sequences in which the NN pair did not contribute to the formation of an extended palindrome. Mirror can bind to an ACAtgTGT sequence (Figure 3.2B) showing that there is no apparent restriction in the nucleotide composition of the central part of the motif. This indicates that the two central nucleotides can be viewed as a spacer acting to separate the two halves of the ACA-TGT palindrome. As suggested by comparing the intensity of the protein-DNA shifts in the experiment shown in figure 3.2 as well as in other experiments (data not shown) a CG pair in the central positions might be slightly more efficient for binding than a TA or TG pair, but this was not tested in direct competition assays.

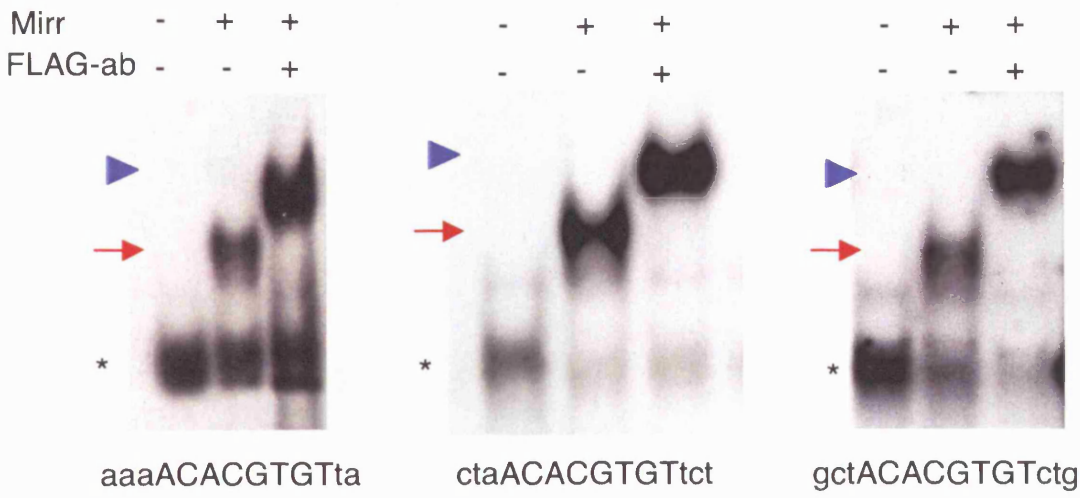


Figure 3.1: Mirror binds a short palindromic sequence ACACGTGT. EMSAs with various oligonucleotides to identify the minimum requirements for binding. The AAAACACGTGTTAA motif was identified through the DNA binding site selection assay. The central part ACACGTGT supports specific binding in different contexts suggesting that the absence of the AT-rich flanking sequences does not compromise Mirror binding. Mutation analysis and competition assays that confirm the specificity of the binding are shown in following figures. In all cases arrows indicate the protein-DNA shift and arrowheads the antibody supershift. The asterisk indicates a non-specific band.

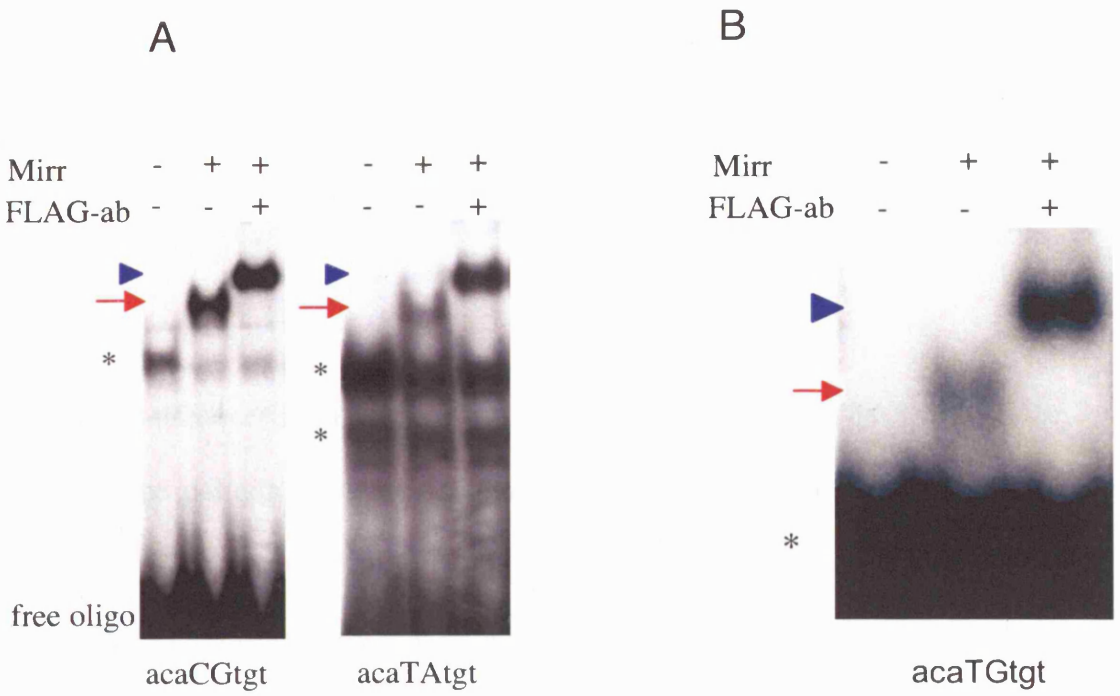


Figure 3.2: The central positions of the ACAcgTGT motif act as a spacer. (A) Changing the two central nucleotides from CG to TA does not abolish binding. Efficiency however (as evinced by comparing the intensity of the protein-DNA shifts) appears to be higher with the CG pair. In both cases equal amounts of labelled probe and equivalent amounts of protein were used. **(B)** The two central nucleotides do not need to form part of the palindrome. Binding also occurs when a TG pair is separating the two half sites ACA and TGT preventing the formation of an extended palindrome as in the case of the ACAC-GTGT or ACAT-ATGT motifs. Arrows indicate the protein-DNA supershift, arrowheads the antibody supershift and the asterisks non-specific bands.

If this hypothesis is true i.e if the function of the central two nucleotides is solely to separate the two half-sites, is there a restriction regarding the length of this spacer sequence? In other words do the two halves of the palindrome need to be separated by two nucleotides?

To address this question I tested Mirror binding to oligonucleotides in which the two halves of the palindrome were separated by a variable number of nucleotides ranging from zero to 12 ($ACA_{n(0\text{-}12)}TGT$). I found that unlike the composition the length of the spacer was actually crucial for the efficiency of the binding as seen by comparing the intensity of the bands in EMSAs where equivalent amounts of protein and labelled probe were used. Sequences with a two-nucleotide spacer bind much better than those with either shorter or longer spacer lengths (Figure 3.3A). Moreover the efficiency of the binding seems to be related to the degree of divergence from the optimal (2nt) spacer length i.e three nucleotides are better than five or none. Specificity is nevertheless maintained as it is possible to detect some weak binding even in cases where the length of the spacer has been increased up to 6 nucleotides and these protein-DNA complexes can still be supershifted by addition of a specific antibody (Figure 3.3B). No binding was detected with spacers longer than 6nt. This suggests that Mirror protein recognizes an ACAnnTGT palindrome, rather than two independent half sites (5'ACA3').

3.1.3. Two direct repeats of the half site (ACAnnACA) are not sufficient for binding

To confirm that a half site (5'ACA3') was not the minimum binding motif I tested Mirror binding to oligonucleotides in which the arrangements of the two halves was changed from inverted to direct repeats, i.e ACAnnACA instead of ACAnnTGT. I was unable to detect any binding to oligonucleotides that carried the direct repeats of the half site (Figure 3.3C).

If Mirror protein recognised and bound the half site (ACA) then increasing the length of the spacer or inverting the orientation of one of the two halves of the palindrome should not have any effect on the affinity of the binding. But Mirror does not bind the ACAnnACA motif nor does it bind oligos in which the length of the spacer has been increased to more than 6nt maintaining the inverted orientation of the half sites (I directly tested 8,11 and 12 nt spacers, data not shown). This indicates that Mirror recognises an ACAnnTGT motif and not a single 5'ACA3' site. To strengthen this hypothesis I also tested Mirror binding to oligonucleotides that only contained the half site (5'ACA3') and could not detect any binding.

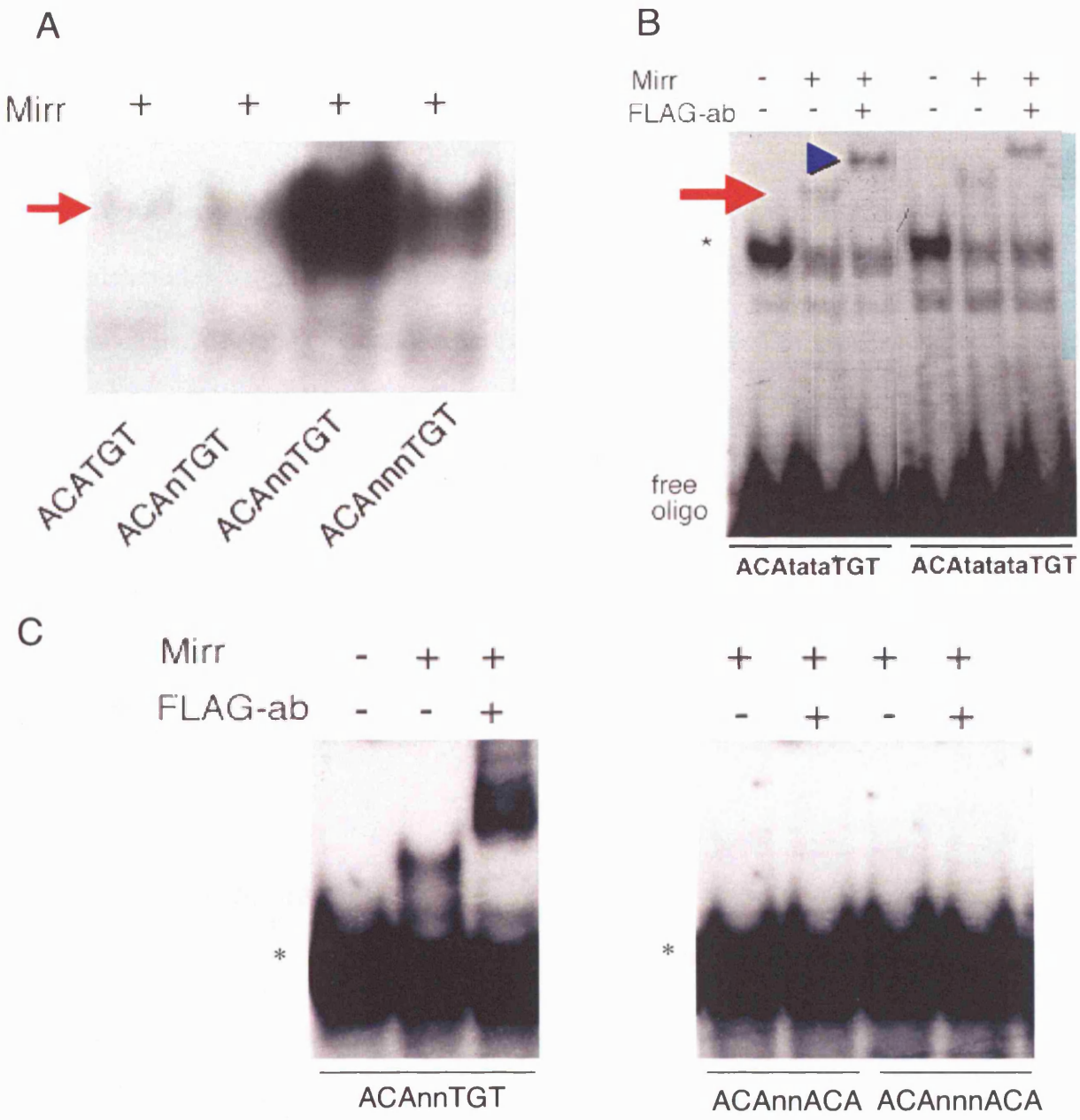


Figure 3.3: The length of the spacer sequence is important for the efficiency of the binding. (A) Varying the length of the spacer sequence from 0 to 3nt has a dramatic impact on binding. Mirror shows a strong preference for a 2-nucleotide spacer and can only weakly bind sequences in which the spacer has been decreased or increased in length. (B) Increasing the length of the spacer to 4 or 6 nt results in further reduction in affinity but maintains the specificity of the binding as shown by the fact that the DNA-protein shifts can still be supershifted by addition of the antibody. (C) Changing the palindromic motif from inverted to direct repeats i.e from ACAnnTGT to ACAnnACA or ACAnnnACA also abolishes Mirror binding.

3.1.4. Single point mutations within each half-site abolish binding

To further validate the specificity of the binding I tested the effect of mutating single nucleotides within each half of the palindrome. I tested two different sets of point mutations. Changing the motif from ACAnnTGT to AtAnnTaT results in almost complete loss of Mirror binding (Figure 3.4). Long exposures of the autoradiography film allowed detection of a much weaker band than the one detected with the intact palindromic motif (data not shown). Introducing point mutations in "non symmetrical" positions at each half site i.e. ACccgTGc also abolishes binding (Figure 3.4). These results strengthen the hypothesis that Mirror specifically recognizes the ACAnnTGT motif.

3.1.5. Competition assays

The specificity of binding can also be assessed by means of competition assays. The principle behind this type of experiment is as follows: if a protein binds specifically to the labelled probe, addition of an excess of unlabelled ("cold") DNA of the same sequence should compete out binding of the protein to the labelled probe. In agreement with this idea, Mirror binding to labelled oligonucleotides that contain the palindromic motif ACAnnTGT is competed out by increasing amounts of unlabelled oligos of the same sequence (Figure 3.5).

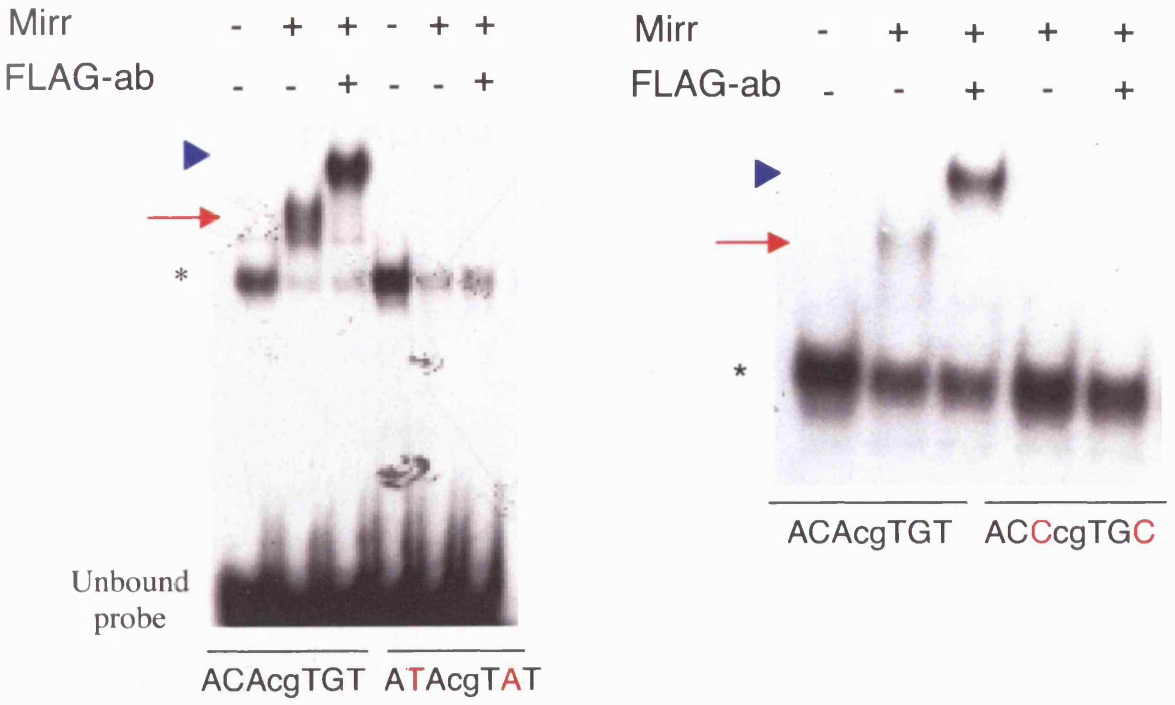


Figure 3.4: Mirror binding on the ACAnnTGT is specific. Point mutations changing the ACAcgTGT core to ATAcgTAT or to ACCcgTGC abolish Mirr binding. In all cases equal amounts of labelled probe were used as seen by comparing the amount of the unbound oligo at the bottom of the left panel. Arrows indicate protein-DNA shifts, arrowheads the antibody supershifts and the asterisks non-specific bands.

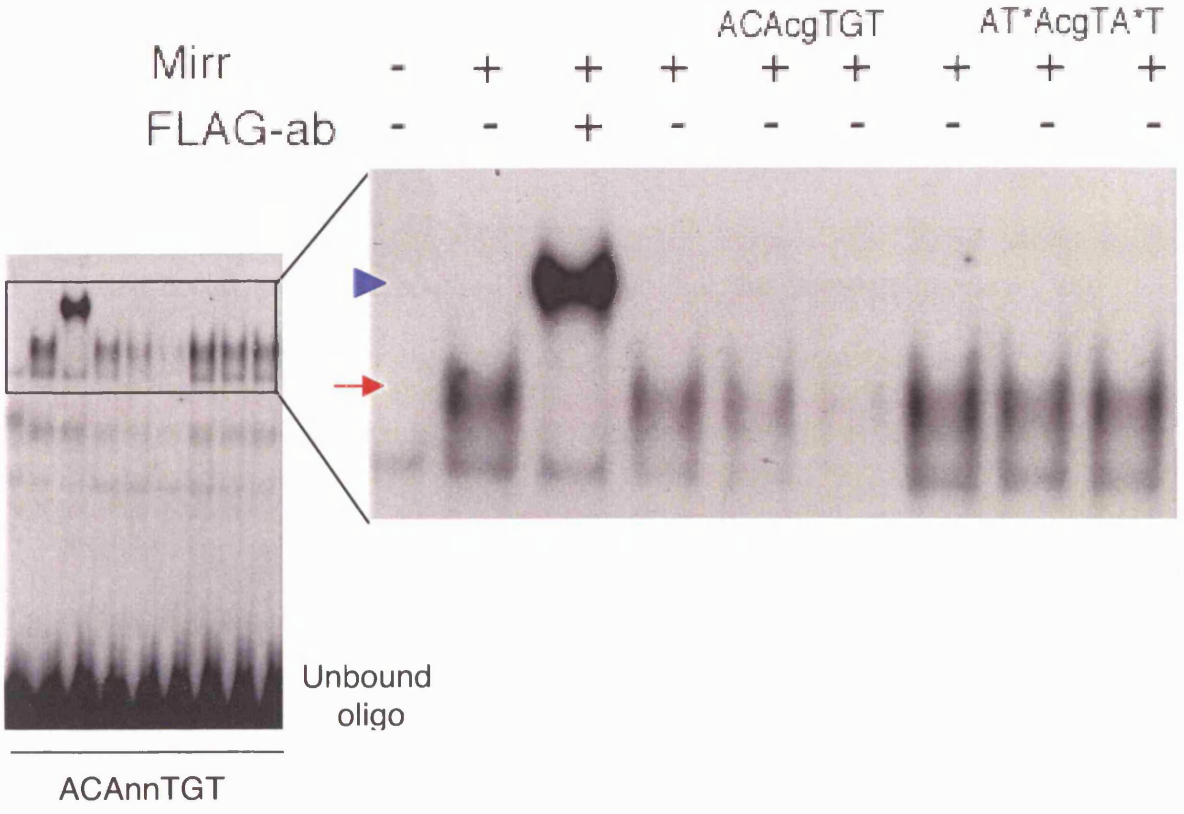


Figure 3.5: Competition studies. On the left panel the whole gel is presented to show that the amount of the labelled oligo (seen at the bottom of the gel) used for the competition analysis was equal amongst all samples. The right panel is a magnification of the boxed region. Binding of Mirror to the ACAnnTGT motif is competed out by increasing amounts of "cold" oligonucleotides of the same sequence but not by oligos carrying the "mutated" ATAnnTAT motif. The arrow indicates the protein-DNA shift and the arrowhead the antibody supershift.

This effect is specific as addition of the same amount of unlabelled oligonucleotides that carry point mutations within each half of the motif (AtAnnTaT) does not affect Mirror binding to the labelled ACAnnTGT probe.

Based on all the above results I concluded that the elements necessary and sufficient for Mirror binding were an ACA and a TGT motif separated by 2 nucleotides. I will therefore be referring to the Mirror binding site as ACAnnTGT.

3.1.6. The E-box consensus motif CAnnTG is not sufficient for binding

Basic Helix-Loop-Helix (bHLH) proteins constitute a large family of transcriptional regulators. The DNA binding basic region (BR) is unstructured when in solution but when bound to DNA it acquires an α-helical conformation that enters the major groove of the DNA (Ma *et al*, 1994). Some bHLH proteins bind DNA as homodimers while others act through heterodimeric complexes (Lassar *et al.*, 1991). Most bHLH dimers recognise the consensus motif CANNTG, known as the E-box, with each monomer binding to a half site (underlined). This motif is very similar to the newly identified Mirror binding site. In fact the E-box is part of the ACAnnTGT motif. I tested Mirror binding to the CAnnTG motif and could not detect any binding (data not shown). This indicates that a complete ACAnnTGT site is necessary for specific Mirror binding to DNA.

3.1.7. Using a bHLH-DNA complex to model Mirror binding to DNA

Despite the degenerate character of the E-box motif, different bHLH proteins display specificity *in vivo*. This is achieved by adopting different BR conformations that allow them to recognise particular subsets of sites (Kophengnavong *et al.*, 2000). Conformational diversity is mediated by particular residues within the BR, which do not come into contact with the DNA. Instead they affect packing of the BR within the major groove and possibly interactions with cofactors. MyoD, a myogenic bHLH protein displays a strong preference for an AACAGCTGTT site, which, based on our studies, should also be recognised by Mirror. MyoD binds this sequence as a homodimer. The structure of the MyoD homodimer bound to the AACAGCTGTT DNA has been resolved by X-ray crystallography (Ma *et al.*, 1994) (Figure 3.6). Using this structure, together with that of PBX1 (the closest to Mirror HD protein for which the structure is available) a theoretical model of a bound-state Mirror-HD was constructed (in collaboration with Bruno Contreras-Moreira). This model suggests that Mirror HD may bind DNA as a homodimer. Although this is just a theoretical approach, it is supported by several facts:

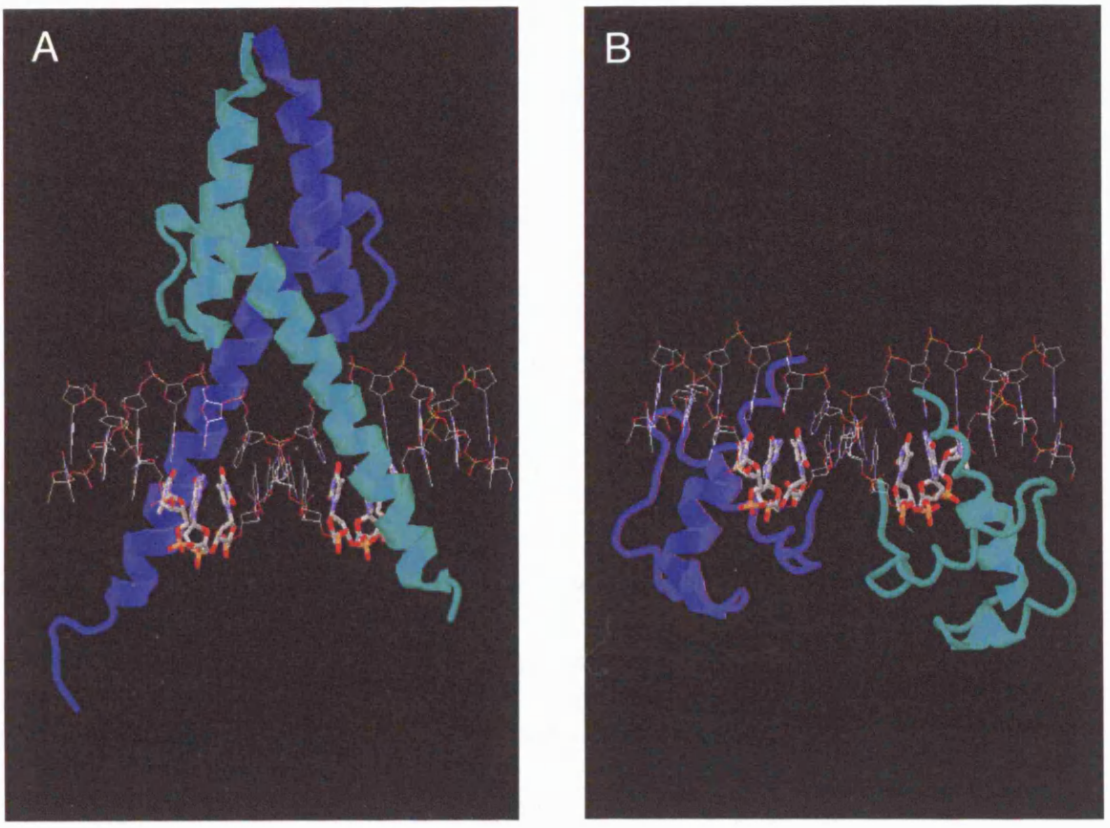
- The helical BR that docks into the major groove is very similar between MyoD and Mirror. Therefore we were able to use the structure of MyoD to dock our Mirror model to the AACAGCTGTT DNA that includes the ACAnnTGT motif.
- Homeodomains do not tend to bend DNA significantly, as seen by comparing HD crystallographic structures stored in the Protein Data Bank. Thus, the

DNA sequence from the MyoD-DNA crystal structure may be used for building of our theoretical model.

- Homeodomains often form dimers. Homodimerisation has been described for several HD transcription factors such as Oct1 (Poellinger and Roeder, 1989), Paired (Wilson *et al.*, 1993), Even-skipped (Hirsch and Aggarwal, 1995), Mix1 (Mead *et al.*, 1996) and others. Heterodimerisation has been demonstrated for HNFα-HNF1β (Mendel *et al.*, 1991), MATα1-MATα2 (Li *et al.*, 1995), Exd-Ubx (Passner *et al.*, 1999) and several others.

3.2. Does Mirror form a homodimer?

The palindromic nature of the Mirror binding site and the modelling of the Mirror HD to this motif suggested that Mirror binding to the DNA might involve formation of a homodimer. This would explain the restrictions in the length of the spacer and the relative orientation of the half sites. In the case of the spacer, it appears that there is some flexibility as to the length that can support dimer formation. As for the orientation, if each Mirror molecule binds a half site (ACA) in the context of an ACAnnTGT palindrome, conformational restraints would not allow binding of a dimer to an ACAnnACA motif. I have tested the possibility of dimer formation in EMSAs and the results will be presented in the following sections.



C

```
1lfu    -KRRNFNKQATEILNEYFYSHLSNPYPSEEAKEELAKKSGITVSQVSNWFGNKRIRYKK
         +|+|      |+ | +     |  ||||+++ |++||++ + | +|||+|| |+| |+||
mirror ARRKNATRETTSTLKAWLNEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKK
                                                          +|| || |
1mdy_H                                                  TTNADRR-KA
```

Figure 3.6: Mirror may bind DNA in a way similar to MyoD. (A) MyoD-DNA complex as determined by X-ray crystallography (based on Ma *et al*, 1994). The MyoD homodimer is bound to the AACAGCTGTT sequence. The TGT DNA motif is highlighted. Note that the blue monomer binds DNA at the back of the DNA molecule whilst the green one binds at the front. (B) Theoretical model of Mirror HD bound to the same sequence. Note that both the green and blue binding helices dock to DNA with similar angles to those in MyoD. The conformation of binding suggests that formation of Mirror homodimers is favourable. (C) Alignment used for model construction, where 1lfu is murine Pbx1 and 1mdy_H is the homologous helical part of MyoD (notice the local similarity). Note that 1lfu is 40% identical in sequence to Mirror, and the alignment has no gaps, making it a very good modelling template. In collaboration with Bruno Contreras Moreira.

3.2.1. Testing homodimer formation in EMSAs

To address the question of homodimer formation I generated an HA-tagged Mirror construct and used it in combination with full-length and partial FLAG-tagged constructs in EMSAs. The principle behind this type of experiment is outlined in Figure 3.7. In the case of homodimer formation one should expect 3 populations of protein-DNA complexes: Some will consist of two HA-tagged molecules, others of two FLAG-tagged proteins and a third population will consist of one HA- and one FLAG-tagged molecule. Since the size of the two proteins is the same, there is no difference in the way these complexes would run on a mobility shift gel. Upon addition of the antibody the situation changes: The third population of homodimers, consisting of proteins carrying different tags can be bound simultaneously by two different antibodies, one recognising the FLAG-tag and one recognising the HA-tag. This higher order complex runs slower in the mobility shift assay, causing a further retardation in the mobility of the free oligo, usually referred to as a super-supershift. If the combination of the two constructs does not support formation of the homodimer, this super-supershift is not detected.

I first expressed HA-Mirror in the *in vitro* transcription/translation system and tested it for binding to the ACAnnTGT palindrome. *In vitro* translated HA-Mirror, as expected, binds DNA with the same efficiency as FLAG-tagged Mirror (data not shown).

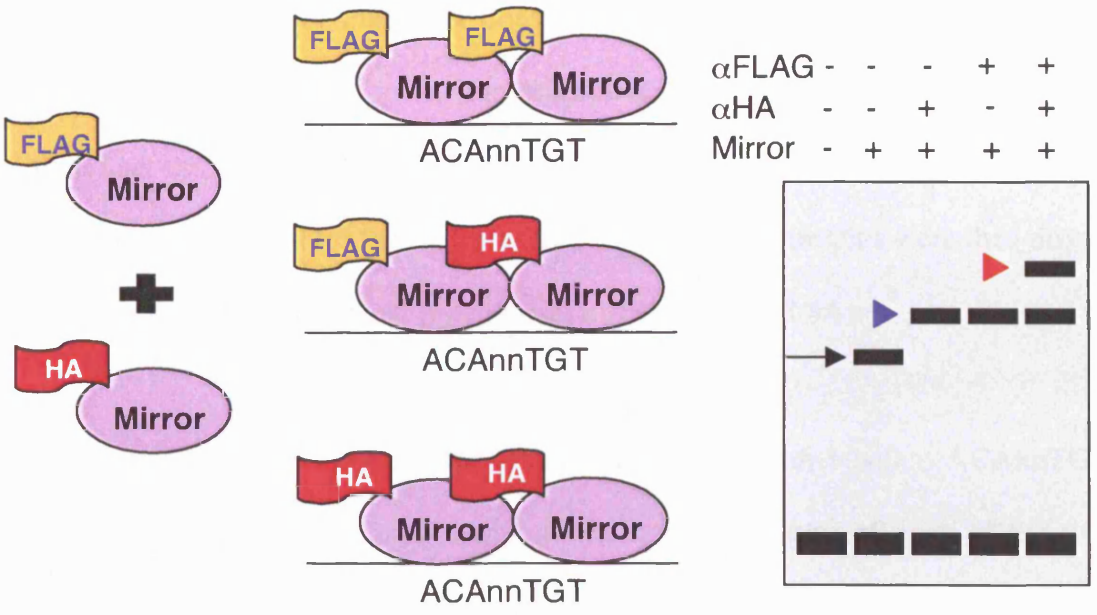


Figure 3.7: Schematic representation of the experimental approach to detect homodimer formation using EMSAs. Two differently tagged Mirror proteins are co-translated in the cell-free system and mixed with the labelled ACAnnTGT probe. In case of homodimer formation one should expect three different types of complexes: one formed by two FLAG-tagged molecules, one formed by two HA-tagged molecules and one containing one molecule of each type. If antibodies against each one of the two different tags are added in the same reaction they would recognise the two tags and cause a super-supershift, i.e a band that will move slower than the single antibody supershift. If there is no dimer formation only one antibody can bind to each protein-DNA complex and the super-supershift is not seen.

When I co-expressed HA- and FLAG-tagged Mirror in the coupled transcription/translation system, expression yields (tested on Western blots) were unequal, due to the fact that both constructs had the same promoter and were competing for the polymerase in the transcription reaction. To overcome this technical problem I set up two separate transcription reactions. The mRNA was quantified and equal amounts were used for the co-translation reaction. The proteins were then mixed with the labelled DNA probe and tested on an EMSA (Figure 3.8A).

When FLAG- and HA-Mirror proteins are mixed with labelled ACAnnTGT probe the protein–DNA complex forms normally and addition of each of the two antibodies causes a specific supershift. Addition of both antibodies gives rise to two bands, one the size of a normal supershift and a second, slower moving band. This shows that antibodies of two different types are binding the same complex indicating that Mirror indeed forms a homodimer.

Interestingly the same result is observed when the two proteins are translated separately and then mixed with the labelled probe, suggesting that dimer formation is a dynamic process and does not depend on simultaneous translation of the two monomers.

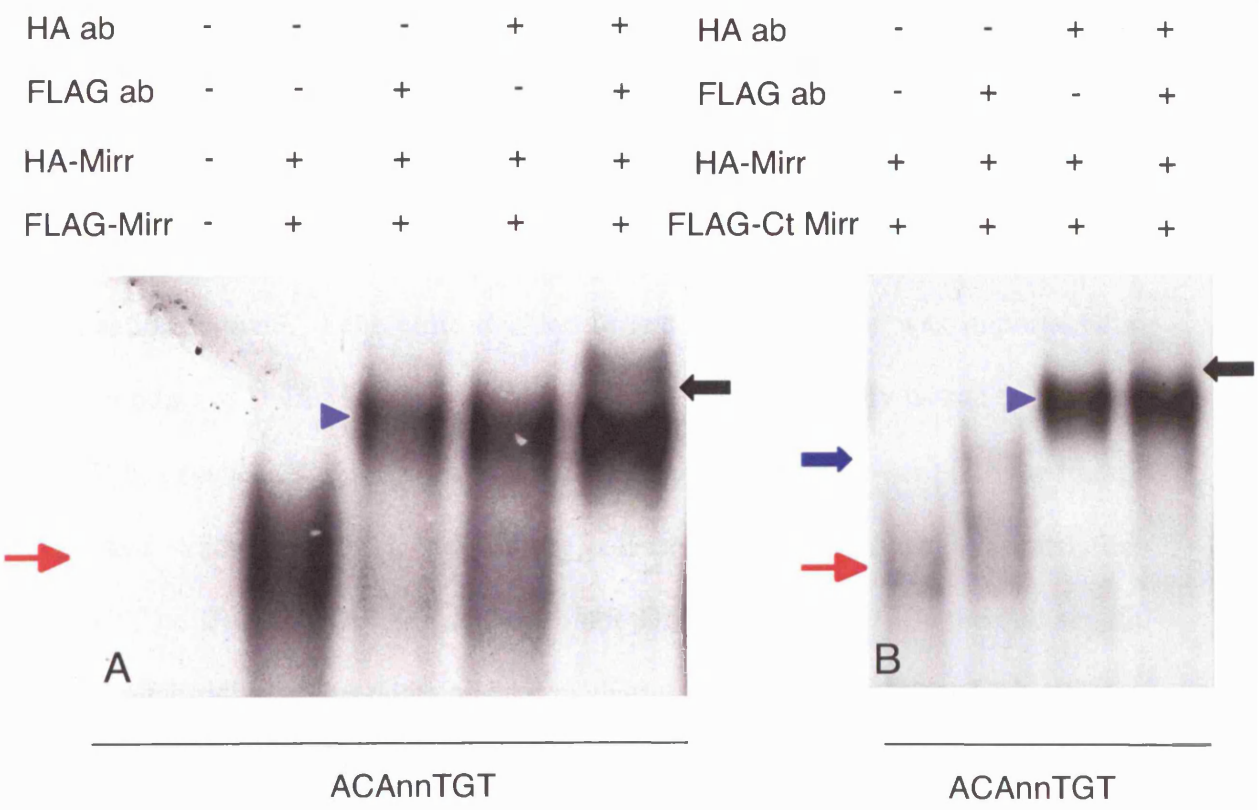


Figure 3.8: Mirror can form homodimers *in vitro*. (A) FLAG-tagged and HA-tagged Mirror are co-translated in the cell-free system and mixed with labelled ACAnnTGT probe. The DNA-protein shift is indicated with a red arrow. Addition of each one of the antibodies against the two different tags results in a supershift, indicated with the blue arrowhead. Addition of both antibodies in the same reaction results in a super-super-shift (thick black arrow), indicating that there are complexes containing both the FLAG and the HA tagged versions of Mirror. (B) If a FLAG-tagged C-terminal Mirror construct is co-translated with full length HA-Mirror, no super-super-shift can be detected (compare thick black arrows in left and right panels) suggesting that elements in the N-terminal region may be necessary for homodimer formation. The FLAG-C-terminal construct gives a lower antibody supershift due to its smaller size (thick blue arrow).

3.2.2. The Iro-box is not essential for binding to DNA.

It has been suggested that the IRO box might be a protein-protein interaction motif and so far has only been found in members of the Iroquois family (Burglin, 1997; McNeill *et al.*, 1997). This made it a good candidate for an Iroquois dimerisation domain. I therefore decided to test if its presence was important for Mirror binding to DNA. I cloned a deletion mutant that specifically lacked the 13aa of the IRO box (generated by site directed mutagenesis by Ina Dahlsveen) into the FLAG vector and synthesized the protein in the cell-free *in vitro* transcription/translation system. The IRO box deletion mutant binds DNA equally well as the full length protein suggesting that this domain is not mediating some homotypic interaction that is important for Mirror binding to DNA (Figure 3.9A).

3.2.3. Residues in the N-terminal region may be involved in homodimer formation

I have also tested partial constructs consisting of the C-terminal end of the molecule, including the HD. Binding of this construct to the DNA was very weak and could only be detected when the antibody was included into the reaction (Figure 3.9B). This "antibody effect" is not surprising as it is known that protein stability is enhanced by formation of higher order complexes. One possible explanation for this result is that binding of the C-terminal construct to the DNA is too weak, making the complex quite unstable. This might mean that some domain in the N-terminal part of the protein is important for proper folding of the molecule so that when this part is missing

binding to DNA is impaired. Alternatively a region in the N-terminal part of the protein may be involved in homodimerisation and this may in turn affect the process of Mirror binding to the DNA. A C-terminal construct that lacks the Homeodomain, as expected, does not bind DNA at all (Figure 3.9B).

Co-translating the HA-tagged full length Mirror and the FLAG-tagged C-terminal construct did not result in the formation of a super-supershift when both antibodies where included in the reaction (Figure 3.8B). This is consistent with the idea that the C-terminal construct may not be sufficient for formation of homodimer and may explain why the protein-DNA shift cannot be detected in the absence of the stabilising effect of the antibody. Further analysis is required to precisely map the domain of homodimerisation. Bandshifts with a N-terminal construct that contains the HD would show whether elements included in this part of the protein are sufficient for formation of the dimer.

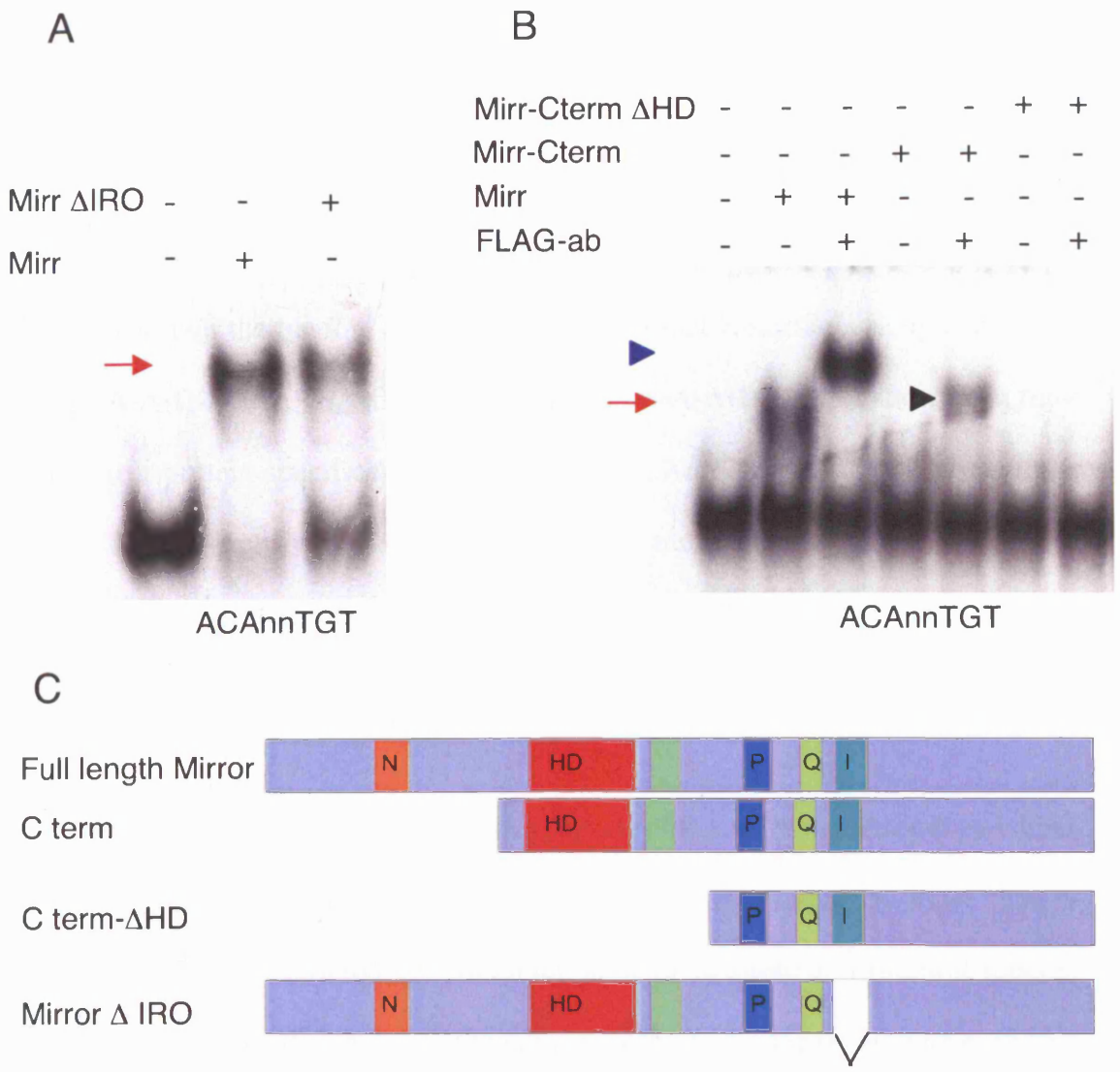


Figure 3.9: Testing various Mirror constructs in EMSAs. (A) A deletion construct lacking the IRO-box binds DNA with comparable affinity as the full-length protein. (B) When a C-terminal Mirror construct was tested on the ACAnnTGT motif only the antibody supershift could be detected (black arrowhead). The DNA-protein shift was not visible. A C-terminal construct that lacks the HD, as expected, cannot bind DNA. The full-length protein is shown for comparison: the red arrow indicates the full-length Mirr-DNA shift and the blue arrowhead the antibody supershift. (C) Schematics of the constructs used for the EMSAs. HD, Homeodomain; I, Iro-box; N, Notch-like EGF motif; P, proline-rich motif; Q, glutamine-rich motif.

3.3. Concluding remarks

In this chapter I have described the minimum requirements for Mirror binding to DNA. Through the binding site selection assay I had identified a long consensus site: **A(g)-A-A(t)-A(t)-A-C-A-C(t)-G(a)-T-G-T-T(a)-A-A(t)**. I have shown that full-length Mirror protein specifically recognises the ACAnnTGT motif included within this site. An AT-rich region at either side of the motif may slightly increase the efficiency of the binding (explaining why this feature has been conserved in the two site selection experiments) but is not essential for recognition.

Mirror protein does not bind to direct repeats of the ACA motif nor does it bind to complete palindromes in which the two halves are separated by more than 6 nucleotides. This observation, in combination with theoretical structural models suggests that Mirror binds DNA as a homodimer with one molecule binding each half of the motif. I have shown in bandshift experiments that Mirror can form homodimers in the presence of DNA and preliminary results suggest that this may require residues in the N-terminal part of the protein. Further experiments are however required to define the exact dimerisation domain.

Chapter 4:

Is the ACAnnTGT motif a universal Iroquois binding site?

4.1. Previous data on Iroquois binding specificities.

Very little has been reported on the binding specificities of any of the Iroquois proteins in flies or in vertebrates. Some downstream targets have been identified, like for example *fringe* in flies (Cho and Choi, 1998; Jordan *et al.*, 2000; Yang *et al.*, 1999) and *Bmp4* in vertebrates (Gomez-Skarmeta *et al.*, 2001) but there is no evidence that these are necessarily direct. Alignment of the Homeodomain of the 3 fly Iroquois with their 6 human homologues shows a perfect conservation within the DNA binding helix implying that binding specificities within this family should be conserved (Figure 4.1)

ara and *caup*, the other two members of the *Drosophila Iroquois* family, have been suggested to control expression of the proneural genes of the *achaete-scute* (*ac-sc*) complex (Gomez-Skarmeta *et al.*, 1996); see also chapter 1. An evolutionary conserved 400bp region located upstream of the *scute* transcriptional start had been previously shown to be important for *ac-sc* expression at the presumptive L3 wing vein (Gomez-Skarmeta *et al.*, 1995). In *Iroquois* deficiencies the L3 proneural cluster is severely reduced or missing suggesting that there might be a direct link between *Iroquois* and *ac-sc* expression.

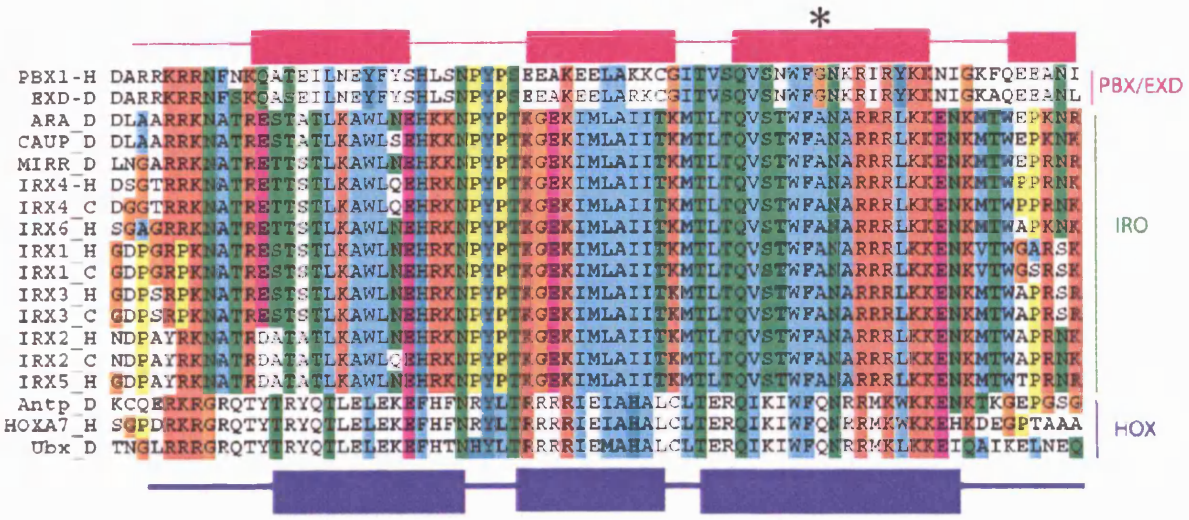


Figure 4.1: Alignment of fly and vertebrate Iroquois Homeodomain sequences. PBX (top) and classic HOX (bottom) Homeodomains (HD) are shown for comparison. Pink boxes correspond to the arrangement of each of the putative α-helices in the PBX and Iroquois HD. Blue boxes show the three α-helices in a classic HD. Iroquois HDs are divergent from classic HDs regarding both the primary sequence and the arrangement of the α-helices. Notice the perfect conservation within the DNA binding helix (asterisk) for all Iroquois members.

To investigate this theory Gómez-Skarmeta and co-workers performed a DNase I protection assay using Ara and the *scute cis*-regulatory region (referred to as the L3 enhancer) to test for direct Ara binding (Gomez-Skarmeta *et al.*, 1996). This assay revealed two contiguous stretches of protected sequence, one of which contained a classic HD binding site (ATTA/TAAT) (Figure 4.2). Based on these results it had been suggested that Ara binds sequences within the L3 enhancer to directly activate expression of genes of the *achaete-scute* complex.

When the L3 enhancer was used in an *in vivo lacZ* reporter assay it was able to drive β-galactosidase expression in the prospective L3 and Twin Sensilla Margin (TSM) territories. When part of the protected sequence was mutated the enhancer failed to drive expression in the normal pattern suggesting that this sequence was involved in direct activation of the *ac-sc* locus (Figure 4.2).

Later work has, however, questioned this idea. This was based on results obtained in *in vivo* reporter assays using Ara fusion constructs. In this type of experiments if a protein acts as a transcriptional activator, fusing its DNA binding domain with the activation domain of a strong transcription activator (like VP16) should result in a fusion-construct that behaves like the wild type protein (i.e that activates transcription). Conversely if the same protein is fused with a repression domain (like the Engrailed repression domain) the fusion construct should act as a repressor, i.e behave opposite to the wild type protein.

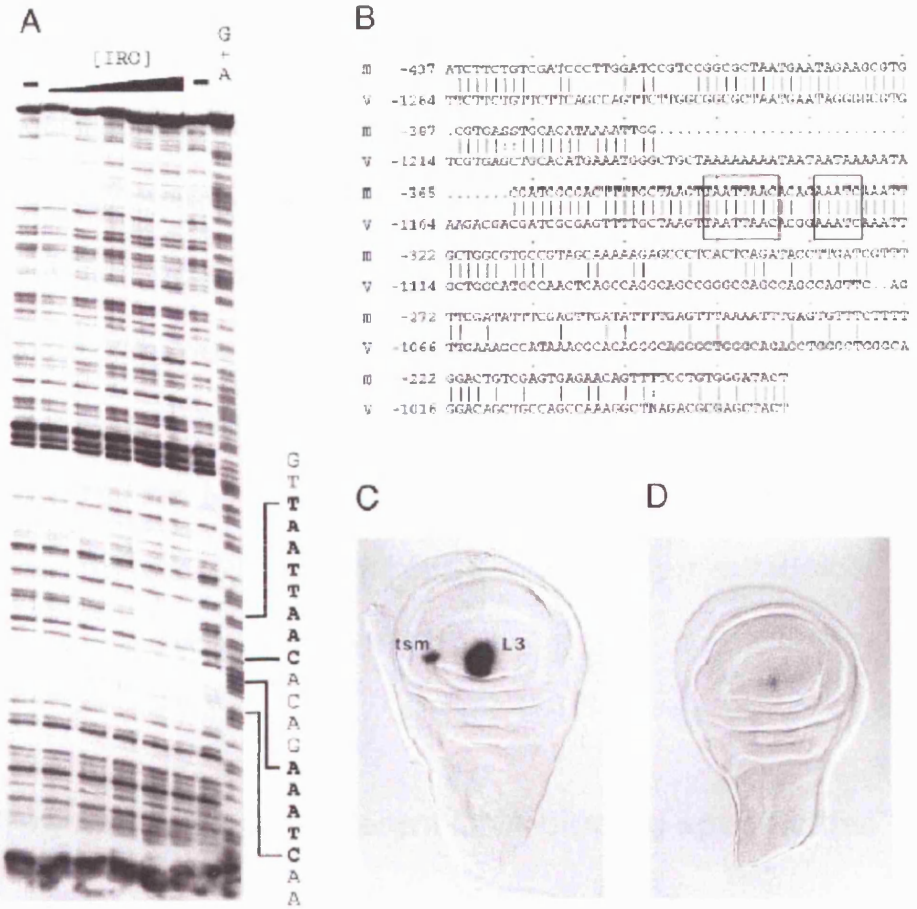


Figure 4.2: Previous data on Iroquois binding specificities. (A) DNase I protection assays using *Drosophila* Araucan had revealed 2 stretches of protected sequences, one of which contained a classic HOX motif. (B) These sequences lie within an evolutionary conserved region as shown in the alignment between the *D. melanogaster* and *D. virilis* genomic DNA upstream of the *scute* gene. (C) Fusing this region to a *lacZ* reporter gene results in expression of β-galactosidase in the region of the L3 vein and the Twin Sensilla Margin (TSM) where the Araucan is expressed. (D) A mutated enhancer in which the TTAATTAA motif was replaced by GCCCCCCC fails to drive β-galactosidase expression in the same regions (taken from Gómez-Skarmeta *et al*, 1996).

Overexpression of a chimeric protein consisting of the Ara HD fused to the Engrailed repressor domain did not behave as a repressor, as one should expect if Ara was acting to activate transcription of the *ac-sc* locus, but resulted in an expansion of the L3 proneural cluster (Cavodeassi *et al.*, 2001). These results were very similar to what was obtained using wild-type Ara suggesting that, contrary to the original idea, Iroquois proteins might in fact act as repressors. In this case their positive effect on the *ac-sc* locus may not be direct and could be mediated by other downstream target(s). Alternatively it could simply relate to lack of proper specification of the wing/notum territory and be a secondary effect.

4.2. Do Mirror and Ara have different DNA binding specificities?

At the time I began the site selection assay the only available data on Iroquois binding specificities were the results of the DNase I protection assay (Gomez-Skarmeta *et al.*, 1996). Based on these results, Iroquois proteins were believed to share the DNA binding specificities of classic HD transcription factors. The site selection however suggested that Mirror had binding specificities different form those of other HD proteins. The difference in the DNA binding affinities between Mirror and classic HOX transcription factors was not surprising based on the divergence of their Homeodomains (Figure 4.1).

Considering the high degree of HD conservation within the Iroquois family one should expect members of the Iroquois family to have similar (or identical) binding specificities. The DNA binding helix is perfectly conserved amongst all members in

flies and in vertebrates (Figure 4.1) and the 3 amino acid differences between the Mirror and Ara HD are of conservative nature.

To test if Mirror and Ara had different binding specificities I tested Mirror binding on the L3 enhancer oligo. Full length Mirror protein synthesized in the *in vitro* translation system binds the L3 enhancer oligo very weakly. In Figure 4.3A a comparison of Mirror binding to the ACAnnTGT site and the L3 enhancer oligo is shown. Both proteins were produced in the *in vitro* system and the yield of the transcription/translation reaction was similar based on Western analysis (data not shown). Equal amounts of labelled probe were used for both binding reactions and it is evident that Mirror binds the ACAnnTGT site with much greater affinity than it binds the L3 enhancer (Figure 4.3A). Competition assays have confirmed this result: Addition of increasing amounts of unlabelled L3 enhancer oligo to a reaction containing Mirror protein and labelled ACAnnTGT probe did not affect Mirror binding to the ACAnnTGT motif. On the contrary the unlabelled ACAnnTGT oligo efficiently competes out Mirror binding to a labelled probe of the same sequence (Figure 4.3B).

I also tested Mirror binding on other classic HD consensus sites, namely these of the Distal Element of the *goosecoid* promoter and the P3 element of the *paired* gene (Germain *et al.*, 2000; Wilson *et al.*, 1993). Both these sequences contain the ATTA/TAAT motif in different backgrounds and in neither case I could detect any Mirror binding (Figure 4.4). From these experiments I concluded that Mirror's preferred binding site was the ACAnnTGT motif and not the previously suggested HOX (ATTA) site.

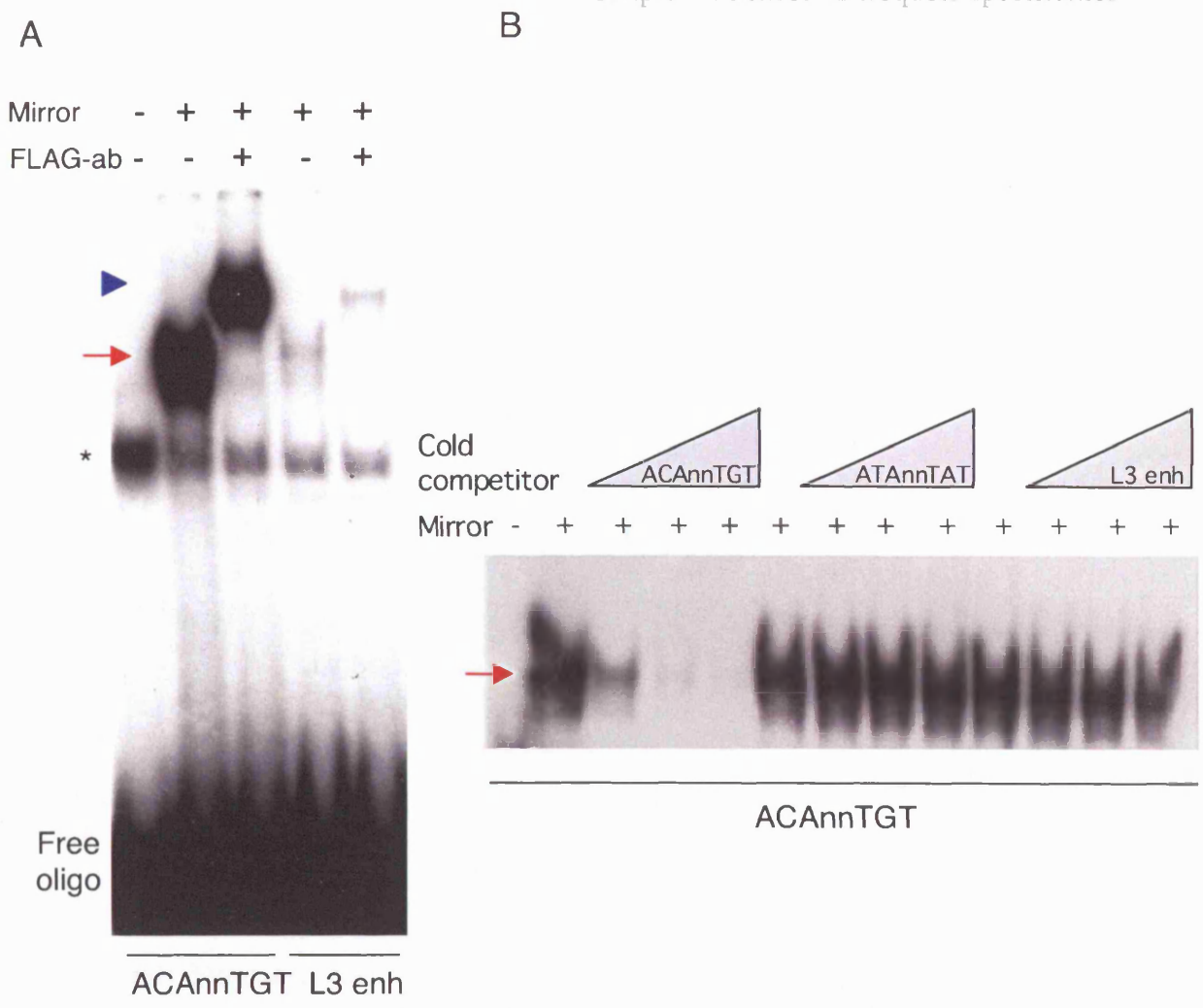


Figure 4.3: Mirror's preferred binding site is the ACAnnTGT motif (A) When equal amounts of labelled probe and protein are used Mirror shows a higher "affinity" for the ACAnnTGT palindrome than for the previously suggested L3 enhancer element that contains a classic ATTA Homeodomain binding site (compare the intensity of the protein-DNA shifts, red arrow) (B) Competition assays. Mirror protein was mixed with labelled ACAnnTGT probe. Increasing amounts of cold oligonucleotides carrying the ACAnnTGT motif compete with the labelled probe for binding to the protein resulting in weaker bands. On the contrary, mutated oligonucleotides (AtAnnTaT) or oligonucleotides that carry the part of the L3 enhancer that was found to be protected in the DNAse I footprinting assay do not result in competition. Note that the L3enh oligo may have some minor effect at high concentrations, which is by no means comparable to the effect of the ACAnnTGT oligo. Arrows indicate protein-DNA shifts and arrowheads antibody supershifts. The asterisk indicates a non-specific band.

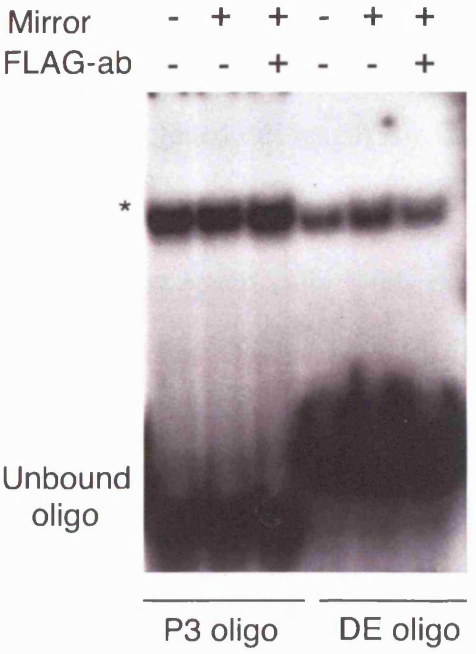


Figure 4.4: Mirror protein does not bind the classic HOX site. Oligonucleotides containing consensus sites for the Paired (P3 oligo) and Goosecoid (DE oligo) Homeodomain transcription factors (classic HOX motif: ATTA) were used in EMSAs with Mirror protein. In none of the two cases binding was detected. The asterisk indicates a non-specific band.

4.3. Ara can bind the ACAnnTGT motif

I then decided to test if any of the other *Drosophila* Iroquois proteins could also bind the ACAnnTGT motif. I cloned full length Ara cDNA into the FLAG vector and expressed it in the *in vitro* system. I performed EMSAs with FLAG-Ara and the ACAnnTGT probe. Full length Ara specifically binds the ACAnnTGT site and the protein-DNA complex can be supershifted with the α-FLAG antibody (Figure 4.5A). Competition assays validate the specificity of the binding as addition of unlabelled oligonucleotides of the same sequence (ACAnnTGT) reduce Ara binding to the labelled ACAnnTGT probe (Figure 4.5B). Ara does not bind the mutated AtAnnTaT motif, as shown both by directly mixing the protein with a labelled AtAnnTaT oligo (data not shown) and by competition assays (Figure 4.5B). Furthermore Ara shows the same preference for a 2nt spacer in the ACAnnTGT motif and efficiency of binding is reduced both with shorter and longer sequences as shown by comparing the intensity of the bands in EMSAs (Figure 4.5A). Nevertheless when equal amounts of probe were mixed with equivalent amount of Mirror and Ara protein (as verified by westerns) Ara binding to the ACAnnTGT palindrome was weaker than that of Mirror. This is a qualitative observation and cannot be validated by competition assays as the proteins in the EMSAs are not radioactively labelled.

Surprisingly, I could not detect any Ara binding on the L3 enhancer sequence. Since Mirror only binds the L3 enhancer oligo with very low affinity this discrepancy may just represent slight differences in binding efficiencies between the two proteins, with Ara binding being below the level of detection.

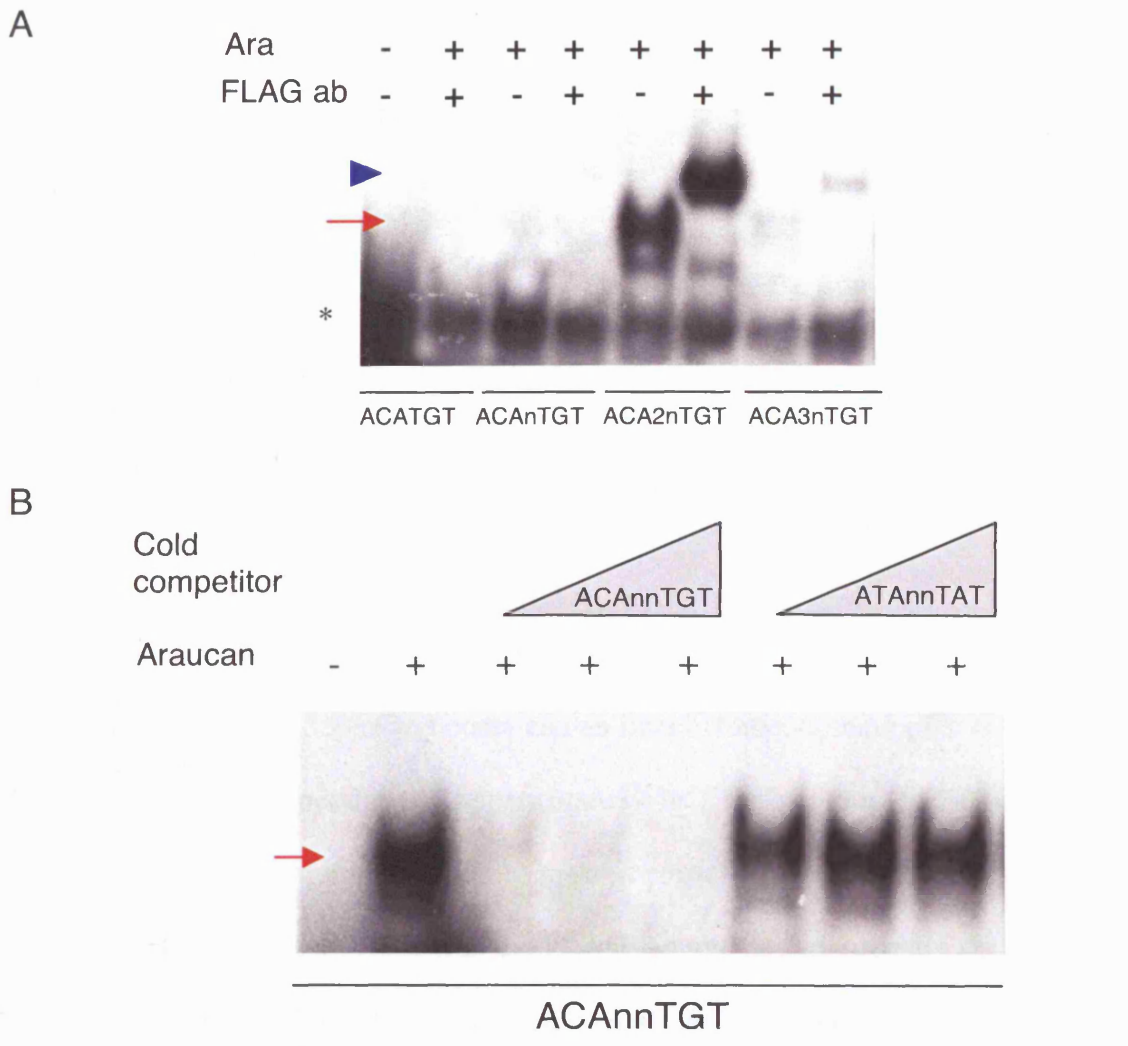


Figure 4.5: Drosophila Araucan binds the ACAnnTGT motif. (A) FLAG-tagged *Drosophila* Araucan binds the ACAnnTGT motif and displays the same preference for a two-nucleotide spacer. (B) Competition analysis with cold oligos showing that binding is specific. Increasing amounts of cold ACAnnTGT oligos efficiently compete out binding of Araucan to the labelled probe. On the contrary equal amounts of a mutated oligo (AtAnnTaT) fail to compete the labelled ACAnnTGT probe. Arrows indicate the protein-DNA shifts and arrowheads the antibody supershifts. The asterisk indicates a non-specific band.

4.4. Comparing full length and HD affinities for DNA.

The fact that Mirror showed some affinity for the L3 enhancer region was puzzling based on our site selection data, as the nature of the consensus motif was distinct from the classic HOX site. Ara can bind the ACAnnTGT site but, in my hands, does not bind the L3 oligo. So how does one explain the results of the DNase I protection assay?

The Ara construct used for the DNase I protection experiment was not full-length. It consisted of ~ 350aa and contained an intact Homeodomain plus N- and C-flanking regions, corresponding to approximately half the protein (full length Ara consists of 716aa). I therefore decided to make a partial Mirror construct consisting solely of the Homeodomain, clone it in the FLAG-tag vector and compare the affinity of the full-length protein to that of the HD on both the ACAnnTGT palindrome and the L3 enhancer motifs.

Expressing the HD constructs in the *in vitro* translation system was not as straight forward as expressing full length Mirror, probably due to the small size of the resulting polypeptide (expected size was approximately 7.5kD). I was never able to detect the HD construct on Western blots using the FLAG antibody. I nevertheless went on to test it on EMSAs, since this assay would also confirm production of the protein by means of the antibody supershift. Interestingly, upon mixing the *in vitro* translated HD-construct with equal amounts of either the ACAnnTGT or the L3 enhancer motifs I could detect comparable binding to both probes (Figure 4.6A).

This result was quite intriguing as it was different to what had been observed for the full-length protein. It implied that although binding to DNA is mediated through the Homeodomain, other regions within the protein are important for the specificity of the binding. This can be viewed as a result of the conformational restrictions imposed by the full-length protein to the intrinsic affinities of the Homeodomain. Alternatively it could be considered as an evolutionary mechanism to achieve specificity in cases where the DNA interacting domain (in this case the Homedomain) has a relaxed specificity for DNA. As shown in the case of Antp and Scr (Furukubo-Tokunaga *et al.*, 1993) (see also introduction-section 1.1.1) changing residues in the N-terminal end of the protein results in distinct preferences for DNA.

I could not however detect any binding of the "HD-alone" construct on the classic ATTA motif of the DE element, the same way I could not detect any binding of the full length protein on the same sequence (Figure 4.6B). This implies that the Iroquois HD maintains some level of specificity and suggests that there may be something else in the L3 enhancer sequences that is recognized by the Iroquois HD.

From these series of experiments one can conclude that in terms of *in vitro* binding full length Mirror protein behaves differently from the isolated HD. Binding preferences may be affected by the overall conformation of the molecule and/or depend on sequences lying outside the HD which are not *per se* involved in protein-DNA interactions.

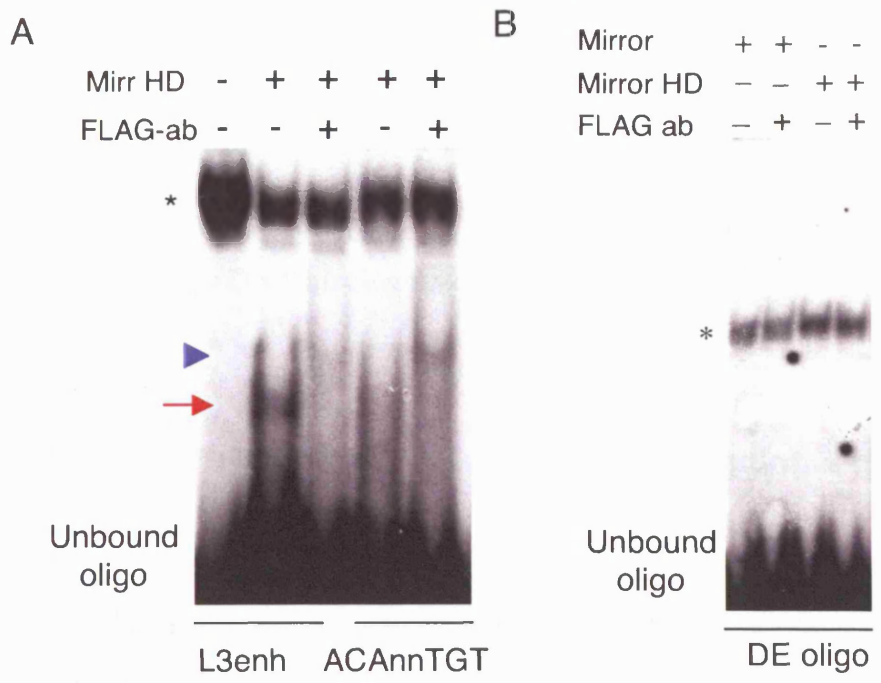


Figure 4.6: Full length and HD-only constructs have distinct binding affinities. (A) The Mirror HD binds comparably to the ACAnnTGT motif and to the L3enhancer (L3enh) oligo. This is different to what was observed with the full length protein (compare to figure 4.3a). (B) Neither the full length nor the HD-only construct show any affinity for the DE oligo that contains the consensus motif for HD transcription factor Goosecoid. In all cases arrows indicate protein-DNA shifts and arrowheads antibody supershifts. The asterisk indicates a non-specific band.

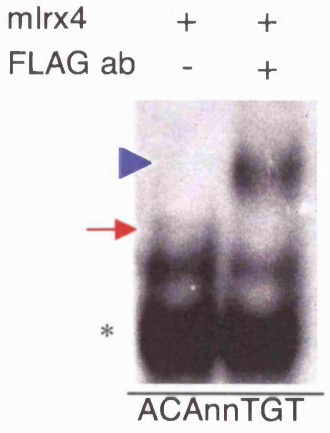


Figure 4.7: Mouse Irx4 can bind the ACAnnTGT motif. EMSA with mouse FLAG-tagged Irx4. The protein-DNA shift obtained is weaker than the one seen using equal amount of probe and Mirror protein (see text). Binding is however specific as seen by mutation analysis (data not shown). The red arrow indicates the DNA protein shift and the arrowhead the antibody supershift. The asterisk indicates a non-specific band.

Deletion mutants that lack the IRO-box bind the ACAnnTGT motif with the same affinity as full length Mirror (see Figure 3.9A), implying that the IRO-box is not involved in defining the DNA binding specificities.

4.5. Is the ACAnnTGT site a universal Iroquois binding site?

One question we were particularly interested in addressing was if the consensus motif we identified by means of the DNA site selection assay was a universal site for Iroquois members. The fact that *Drosophila* Ara can bind the ACAnnTGT motif suggested that the motif we identified through the site selection may be of a more general nature and could be recognised by other non-*Drosophila* Iroquois members. To test this hypothesis I subcloned the cDNA from a vertebrate homologue, mouse Irx4, into the FLAG vector and tested it on EMSAs. Mouse Irx4 produced in the *in vitro* system binds the ACAnnTGT motif (Figure 4.7). However using equal amounts of labelled probe and protein (the latter was tested in Western blots) mIrx4 binding to the ACAnnTGT motif seems to be weaker than that of Mirror (and of Ara). Binding is specific as shown by the fact that Irx4 does not bind the mutated palindrome (AtAnnTaT) (data not shown). Consistent to what we saw for Mirror, neither Ara nor Irx4 can bind the classic HOX motif (TAAT). Furthermore and similar to Ara, Irx4 does not bind the L3 enhancer oligo (data not shown).

Together these results suggest that the novel binding motif identified for Mirror protein may be a generic/universal motif for Iroquois proteins and I will therefore refer to it as an Iroquois Binding Site or IBS. We cannot, however, exclude the possibility

that other Iroquois proteins have slightly different preferences. Both Ara and Irx4 show a weaker binding to the ACAnnTGT motif than Mirror and it is possible that their optimal binding site, although related, may be slightly different. We thought that spacing of the two halves of the palindrome may be different for different Iroquois members. But this is not the case, as Ara behaves like Mirror when tested on palindromes with varying spacer lengths (Figure 4.5A). It would be interesting to obtain *in vitro* data for the binding specificities of other Iroquois members. This sort of information would be very useful not only to understand the nature of Iroquois binding to DNA but also, in more general terms, to get an insight on HD conservation and evolution within members of the same family.

4.6. Mirror can form heterodimers with Ara on the ACAnnTGT site.

The fact that Mirror and Ara can bind the same site together with the observation that Mirror can form homodimers *in vitro* suggested that Mirror might also form heterodimers with Ara (and/or Caup). To test this possibility I decided to co-translate the two proteins and look for heterodimer formation by means of an EMSA. Due to the fact that both the HA-Mirror and the FLAG-Ara constructs had a T7 promoter for *in vitro* transcription I set up two separate transcription reactions to make sure I was not affecting the stoichiometry of dimer formation (see also chapter 3 for HA- and FLAG-Mirror). I then used equal amount of the two mRNAs in the same *in vitro* translation reaction. Upon mixing the labelled ACAnnTGT probe with FLAG-Ara and HA-Mirror a protein-DNA shift could be detected. Addition of each one of the antibodies caused a specific supershift showing that each one of the two proteins was binding DNA. When both antibodies were added to the same binding reaction

they caused formation of a super-supershift indicating that both Mirror and Ara were part of the same protein-DNA complex (Figure 4.8).

This observation suggested that other Iroquois proteins may also form heterodimers. This could be an additional mechanism to increase specificity as it is possible that heterodimers have slightly different preferences for binding to the DNA than homodimers e.g. in terms of spacing or flanking sequences. As *mirror* and the other two *Iroquois* genes have a significant percentage of overlapping expression patterns it is possible that in some tissues or developmental stages they may act cooperatively through heterodimers to control expression of target genes.

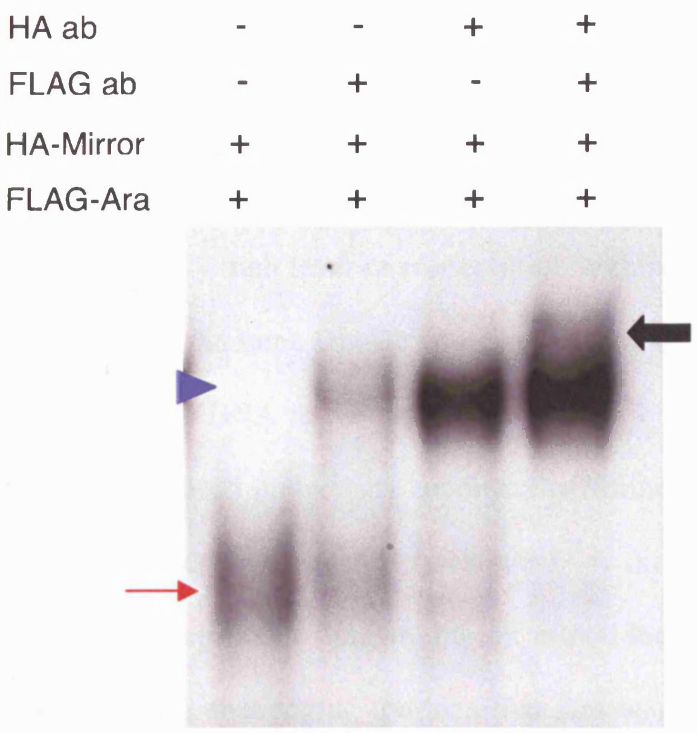


Figure 4.8: Mirror can form heterodimers with Araucan. EMSAs performed with HA-Mirror and FLAG-Araucan have shown that the two proteins when co-expressed in the *in vitro* translation system can form heterodimers upon binding to the DNA. The protein DNA shift is indicated by the red arrow. The supershifts caused by each one of the two antibodies are indicated by the blue arrowhead. A super-super-shift (thick black arrow) is formed when both proteins and both antibodies are present in the same binding reaction suggesting that the two proteins are part of the same complex. See also Figure 3.8 for formation of a Mirror-Mirror homodimer and a N-term Mirror deletion which fails to form the super-supershift.

4.7. Concluding remarks

In this chapter I present a comparative study on the affinity of Iroquois proteins for DNA. There are three *Drosophila Iroquois* genes while in mice and humans there are six. All genes show a very high level of conservation within the Homeodomain, implying that they may have the same DNA binding specificities. Previous data had suggested that they may bind DNA in the same way as classic Homeodomains do. This idea was based on a DNase I protection experiment performed with a partial Ara HD construct. These data were subsequently questioned. A fusion construct of the Ara HD with the Engrailed repression domain did not mimic the effect of wild type Ara (Cavodeassi *et al*, 2001) challenging the idea that Ara was directly activating transcription of the gene *scute*.

I have shown that *Drosophila* Ara and a vertebrate homologue, mouse Irx4 can bind the ACAnnTGT palindrome suggesting that this site is a generic Iroquois Binding Site (IBS).

The Mirror HD alone cannot discriminate between the ACAnnTGT site and the HOX-like (ATTA) motif previously suggested to be bound by Ara in the context of the L3 enhancer. I speculated that the reason this motif was identified by means of the DNAse protection assay was that a partial Ara construct was used for the experiment. When I looked within the 400bp fragment that was used for the DNase footprinting experiment I could not find any complete ACAnnTGT motifs, which agrees with the hypothesis that the effect of Iroquois proteins on the genes of the *ac-sc* complex is not direct (Cavodeassi *et al.*, 2001). There are, however, two overlapping

half sites of the Mirror binding motif (ACACA) immediately downstream of the TAAT motif (TAATTAACACAGAAATC: underlined are the protected stretches of sequence showing that the ACA motif is partially protected). Note that this is reminiscent of an ACA half site in the context of the AT-rich flanking sequences as seen in the site selection consensus. Mutating part of the protected ACA to AtA results in loss of binding for both the full length and the HD construct (data not shown). However, the same is true for mutations within the protected ATTA sequence. These observations together with fact that binding is in general too weak to obtain reliable competition results using labelled L3 probe only allow a speculative interpretation of these results.

One of the putative direct targets for vertebrate Iroquois proteins is *Bmp-4*. This has been suggested in several cases for Iroquois members from different species (Gomez-Skarmeta *et al.*, 2001; Goriely *et al.*, 1999; Kudoh and Dawid, 2001). Gómez-Skarmeta and co-workers showed that a partial Xiro1 construct could bind directly to an enhancer element of the Xenopus *Bmp4* gene that contains HOX-like sites. They also showed that the partial Ara HD construct previously used for the DNase I protection assay can also bind the *Bmp4* element. There are no IBSs within this *Bmp4* genomic fragment, however there are several half sites (ACA and TGT in various orientations and with various spacing). In a series of fusion experiments with the VP16 activation and the Engrailed repression domain Gómez-Skarmeta and co-workers showed that Xiro1 represses the expression of Bmp-4. It is not clear whether these results are representative of the effects of the full length protein but, based on our observations, when partial constructs are used it should be kept in mind that their binding specificities may be different from those of the full length protein. For this reason the results of these experiments should be interpreted with caution.

I have shown that, in addition to homodimers, Mirror can also form heterodimers with Ara, at least upon binding to the ACAnnTGT motif. Members of the NK2 class of HD transcription factors have also been shown to form homodimers as well as heterodimers with other members of the same family (Kasahara *et al.*, 2001). Paired/Pax class of Homeodomains also bind cooperatively to palindromic DNA as homo- and heterodimers (Wilson *et al.*, 1993). In both cases (NK2 and Paired class) dimer formation increases the affinity of the protein-DNA interaction but monomeric binding (on the half site) is also possible. In the case of the Paired/Pax class different members require different lengths of spacer separating the two half-sites. It will be interesting to test whether changing the length of the spacer also has an effect on heterodimer formation in the case of Iroquois proteins.

Chapter 5:

In vivo validation of the novel Iroquois Binding Site (IBS)

In the last two chapters I presented work done to characterise a newly identified binding site for Mirror and other members of the Iroquois family of transcription factors. This site is distinct from the classic Homeodomain consensus motif and can be recognised by members of the Iroquois family both in vertebrates and invertebrates. Due to the nature of the selection assay this motif was identified for its high affinity for Mirror protein *in vitro* and therefore may not necessarily be functional *in vivo*. In this chapter I will present evidence that Mirror can recognise the ACAnnTGT motif *in vivo* and that reporter constructs that carry a tetramer of this site can be transcriptionally regulated by Mirror in transgenic flies.

5.1. Mirror produced in S2 cells binds the ACAnnTGT site

The *in vitro* site selection assay relies on the affinity of a given protein for naked DNA and does not take into account parameters such as *cis*-element morphology and requirement for trans-acting factors that may modify DNA binding specificities. It is possible that transcription factors, showing a strong affinity for a particular site *in vitro,* exhibit different preferences *in vivo* due to formation of complexes with other transcription factors that may or may not bind DNA themselves. It has already been shown that HD transcription factors acquire different binding specificities upon heterodimerisation with members of the PBX/Exd family of TALE

transcription factors (Chang *et al.*, 1996). *In vivo* binding may, in other words, discriminate between sites that *in vitro* would be bound with equal affinity.

The protein used for the site selection and the characterisation of the site was produced using an *in vitro* transcription-translation system. For this reason one would expect a possible lack of potential co-factors and possibly some of the post-translational modifications that normally occur in the cell. Mirror has previously been shown to be phosphorylated in S2 cells (Trevor Littlewood and Helen McNeill, unpublished results). To test if Mirror can bind the selected site in cells I transfected *Drosophila* S2 cells with a FLAG-Mirror construct under the control of a metallothionein-inducible promoter and generated stable lines that, as shown in Western blots, express Mirror when induced with $CuSO_4$ (Figure 5.1A). Mirror expressed in S2 cells runs slightly slower than *in vitro* translated Mirror indicating that phosphorylation or other modifications that occur in the cell may not occur *in vitro*.

Nuclear extracts of Mirror-expressing S2 cells were used in EMSAs with the ACAnnTGT probe. Under the conditions used for *in vitro* translated Mirror no binding was detected. I tested various parameters such as the concentration of salt, non-specific competitor (polydI-polydC) and detergent but none of these had any effect (data not shown). However increasing the Mg^{2+} concentration from 3mM to 6mM allows Mirror binding to the ACAnnTGT probe, showing that Mirror produced in cells can bind the ACAnnTGT motif (Figure 5.1B).

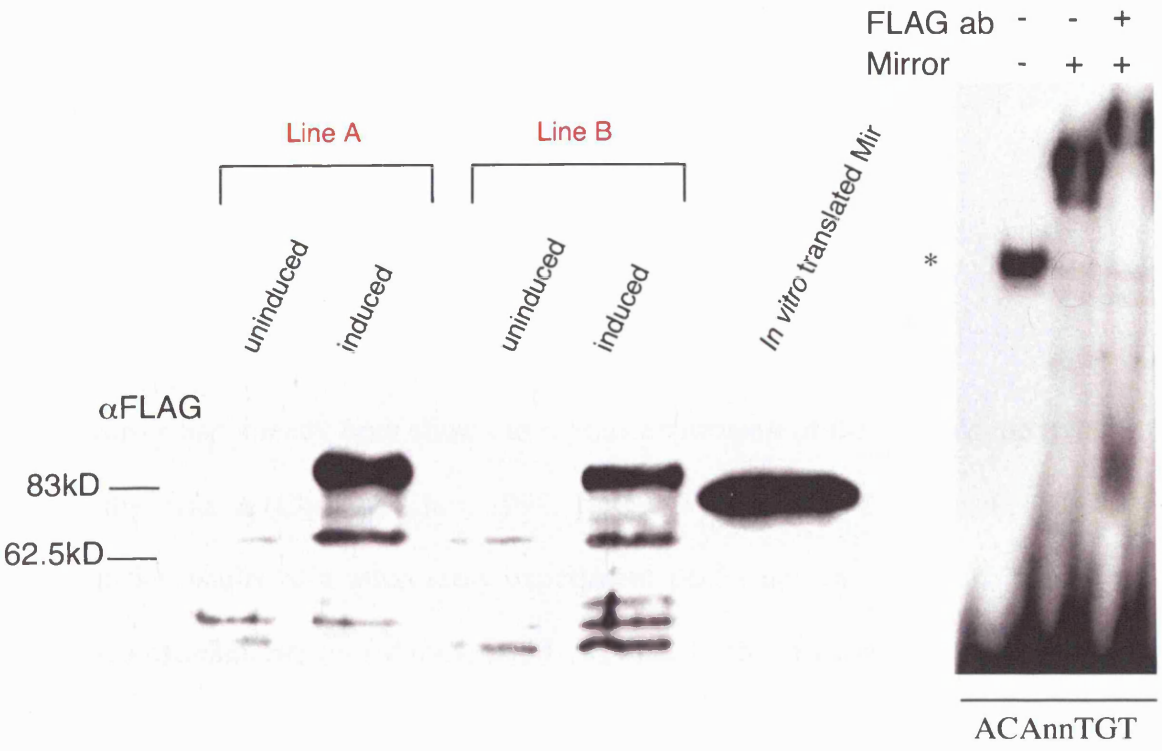


Figure 5.1: Mirror protein produced in S2 cells can bind the ACAnnTGT motif. (A) Two stable S2 cell lines expressing Mirror under the control of a metallothionein inducible promoter were generated. Addition of $CuSO_4$ into the growing medium leads to induction of Mirror expression as seen by immunoblotting with the α-FLAG antibody. (NB Mirror is not endogenously expressed in S2 cells). *In vitro* translated Mirror is shown for comparison. Mirror produced in S2 cells runs slightly slower than *in vitro* translated Mirror in the 10% gel, implying that Mirror may be post-translationally modified. (B) Nuclear extracts from line A were used in EMSAs with labelled ACAnnTGT probe showing that Mirror produced in cells can recognise the same motif. The same was observed with extracts from line 2 (data not shown).

This observation suggested that the IBS could be functional *in vivo*. We decided to test this hypothesis by generating *in vivo* transgenic reporter lines.

5.2. Making of IBS-*lacZ* transgenic lines.

Mirror had already been shown to repress expression of the *fringe* gene in the eye and the ovaries (Cho and Choi, 1998; Jordan *et al.*, 2000; Yang *et al.*, 1999). Based on the results of a microarray experiment performed in our lab to identify Mirror downstream targets (Mohns, 2003) we had further evidence that Mirror in many cases acts as a transcriptional repressor. For this reason we decided to test Mirror for the ability to mediate transcriptional control through the IBS in transgenic lines using a system that would allow us to detect repression of transcription. We used the pHZ50PL-Gbe vector (Jennings *et al.*, 1999) that carries three sites for the transcriptional activator Grainyhead upstream of the *lacZ* gene (Figure 5.2A). *LacZ* reporter constructs that carry these Grainyhead binding elements (Gbe) drive ubiquitous expression of β-galactosidase in all imaginal discs (Figure 5.2B and C).

I cloned a 110bp oligonucleotide carrying 4 repeats of the ACAcgTGT motif into the pHZ50PL-Gbe vector (Figure 5.4A, see section 7.1.5.5 for the complete sequence of the insert). The construct carrying the IBS tetramer in the context of the Grainyhead elements was then injected into embryos to generate transgenic lines (injections performed by Terence Gilbank and Steven Murray).

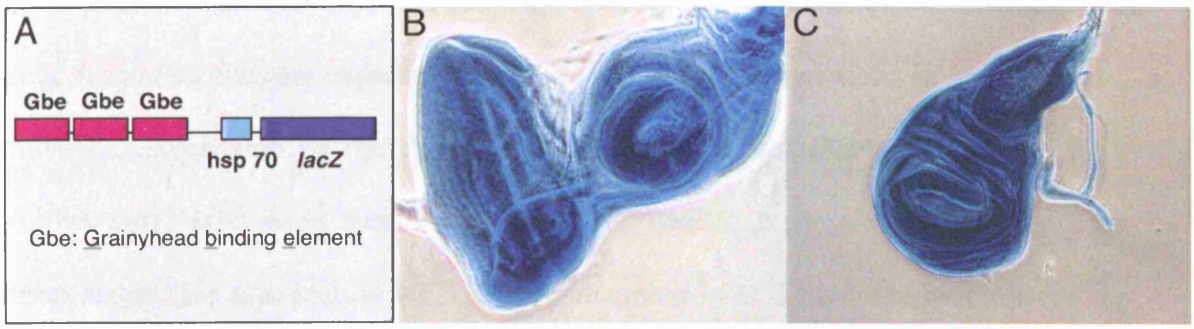


Figure 5.2: The Gbe-LacZ construct drives ubiquitous expression of β-galactosidase in imaginal discs. (A) Diagram of the vector used for the *in vivo* reporter analysis. Three copies of the Grainyhead binding element (Gbe) have been inserted upstream of the *lacZ* gene (Jennings *et al,* 1999). (B) X-gal staining of an eye–antennal disc from the Gbe-lacZ line showing that expression of β-galactosidase is ubiquitous. (C) A wing disc from the same line also exhibits ubiquitous β-galactosidase expression.

We obtained 2 independent lines, corresponding to two separate insertion events in the germ line. The two reporter lines showed the same expression pattern although one of the two lines gave slightly stronger staining than the other.

I looked for β-galactosidase expression in the eye and wing imaginal discs of the Gbe-IBS-*lacZ* lines where *Iroquois* expression has been extensively studied. All three *Iroquois* genes are expressed in the dorsal half of the eye disc. In the wing disc, however, expression of the three *Iroquois* genes is not identical (Figure 5.3). Furthermore, expression patterns are slightly different during the second and third instar stages (see also section 1.3.3.2.). The expression of β-galactosidase in the wing disc of the Gbe-IBS-*lacZ* flies is shown in Figure 5.4B. X-gal staining revealed an extended repression in β-galactosidase expression compared to the Gbe-lacZ line (compare to Figure 5.2C). Interestingly, the repression domain includes the notum region where the expression of the three *Iroquois* genes overlaps. We decided, however, to focus our studies on the eye disc where *Iroquois* expression is much more localised making it easier to analyse its effects on the reporter construct.

In the eye disc, introducing the IBS tetramer in the Gbe background results in strong repression in expression of β-galactosidase in the dorsal half of the disc. Expression is maintained ventrally, stronger at the pole and weaker towards the midline. Staining was mainly restricted to the anterior part of the disc, ahead of the morphogenetic furrow (Figure 5.5B and C).

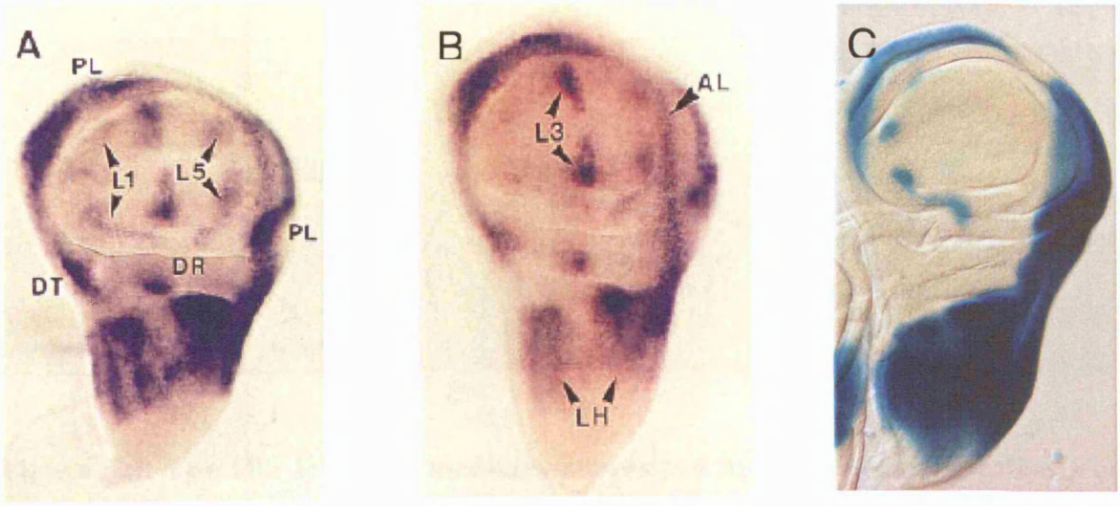


Figure 5.3: Expression of the three *Iroquois* genes in the *Drosophila* wing disc. (A and B) Expression of *araucan* and *caupolican* as revealed by *in situ* hybridisation with full length RNA probes. Expression occurs in the lateral heminotum (LH), distal tegula (DT), pleura (PL), alula (AL) veins L3 and L5 and proximal vein L1 region. Taken from Gómez-Skarmeta et al, 1996 (C) X-gal staining of a *mirror*-lacZ line showing expression in the lateral heminotum, the alula and pleura regions. Note that *mirror* does not seem to be expressed in the presumptive L3 and L5 vein.

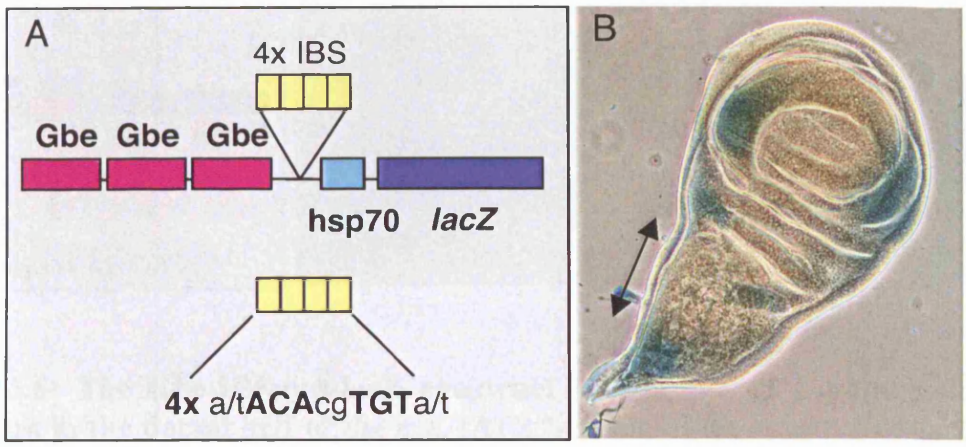


Figure 5.4: Expression of the IBS-lacZ in the *Drosophila* wing disc. (A) Schematic of the Gbe-lacZ vector with the IBS tetramer introduced between the GBEs and the reporter gene. (B) Expression of the Gbe-IBS-lacZ construct in the wing disc. Strong repression is observed. Note that this includes the notum region (double-headed arrow) where the expression of the three *Iroquois* genes is overlapping (see also figure 5.3).

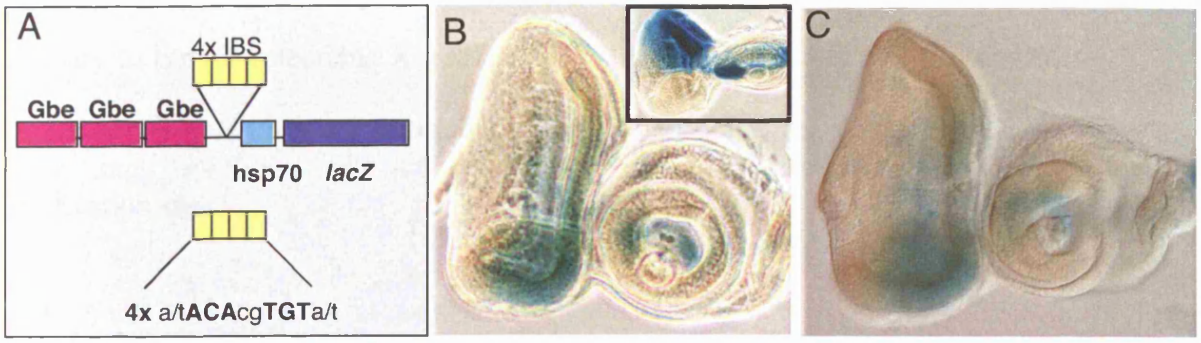


Figure 5.5: The IBS tetramer mediates repression in the eye imaginal disc. (A) The structure of the Gbe-IBS-lacZ. (B and C) Two independent Gbe-IBS-lacZ lines showing the same pattern of β-galactosidase expression by activity (X-gal) staining. Repression is observed in the dorsal half of the disc where *mirror* and the rest of the *iroquois* genes are expressed (compare to inset in B: X-gal staining of a *mirror*-lacZ line). Note that there are regions in the disc where repression cannot be attributed to the Iroquois (e.g in the ventral posterior part of the disc). Discs are shown with dorsal to the top and posterior to the left.

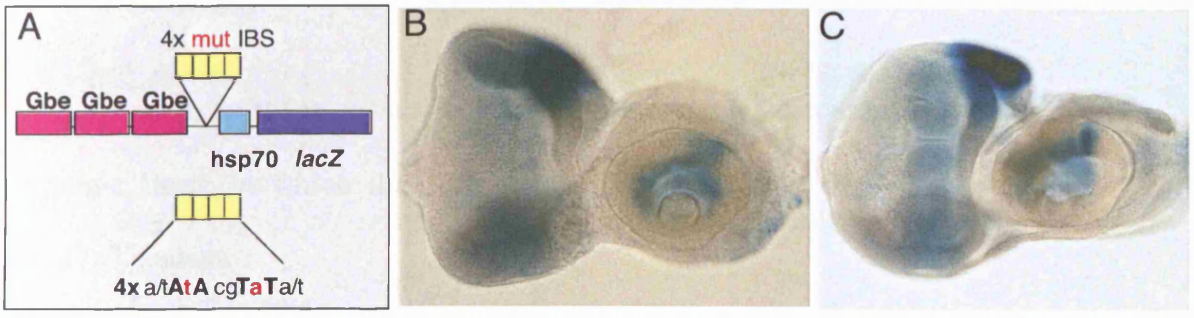


Figure 5.6: The Gbe-IBSmut-lacZ construct causes loss of β-galactosidase repression in the dorsal half of the eye. (A) Schematic of the structure of the Gbe-IBSmut-lacZ construct. (B and C) The two independent lines have similar expression patterns. X-gal staining reveals that there is a dorsal de-repression in the β-galactosidase pattern that coincides with the region where the Iroquois should be acting (see inset in figure 5.5B). Discs are shown with dorsal to the top and posterior to the left.

In both Gbe-IBS*lacZ* lines presence of two copies of the transgene was necessary to obtain detectable X-gal staining. I could not obtain any staining using various anti-β-galactosidase antibodies, even when applying a biotin-streptavidin amplification step.

The repression described above is strong in the dorsal half of the disc where *mirror* and the other *Iroquois* genes are expressed at their highest levels. Clearly there is also repression in other parts of the disc (such as the region behind the morphogenetic furrow) implying there might be other proteins acting on this element. In addition, it could be argued that increasing the distance between the Grainyhead elements and the *lacZ* gene by introducing the IBS tetramer may also cause a reduction in the levels of transcription activation mediated by the GBE.

To address these issues I decided to test β-galactosidase expression in transgenic lines, in which the IBS would be replaced by the binding-deficient AtAnnTaT mutant.

5.3. Making of IBS-mutant-*lacZ* reporter lines

To generate the IBS mutant lines I used an oligonucleotide that carried 4 repeats of the IBS with single point mutations within each half site (i.e AtAcgTaT) in the same context and orientation as the IBS sites used for the previous experiment (Figure 5.6A). The oligo was cloned into the Gbe-*lacZ* vector using the same restriction sites and the construct was injected in embryos to generate transgenic lines (injections performed by Genetic Services). This arrangement allowed me to assess

the positional effects mentioned above. We obtained two independent lines that gave similar results.

The expression pattern of the IBS mutant lines is shown in Figure 5.6 (B and C). Unlike what was observed with the Gbe-IBS-*lacZ* reporter the Gbe-IBSmut-*lacZ* construct does not mediate β-galactosidase repression in the dorsal part of the disc. This de-repression greatly overlaps with the *mirror/Iroquois* expression pattern at the dorsal anterior part of the disc. Staining in other parts of the disc remains unaffected. Thus, changing the insert sequence from ACAcgTGT to AtAcgTaT in the same Gbe-*lacZ* background results in a loss of the dorsal repression of β-galactosidase. This result demonstrates that an intact ACAnnTGT site is necessary for dorsal repression. It also confirms that the repression observed with the IBS-*lacZ* construct cannot be solely attributed to a positional effect, since the two constructs have exactly the same length. It does not, however, conclusively prove that repression is due to Mirror.

Another way of interpreting the above results would be to attribute the dorsal expression of β-galactosidase in the IBS-mutant lines to the presence, in the mutant construct, of a binding site for a transcription activator expressed in the dorsal half of the eye disc. The Wg pathway effector *pangolin* (dTCF) is a candidate for this role but there are no Pangolin binding sites (CCTTTGATCTT, (Lee and Frasch, 2000) within the IBSmutant construct. The TALE HD protein Homothorax (Hth) has recently been shown to be expressed in a narrow stripe at the Dorsal Rim Area of the eye (Wernet *et al.*, 2003). There are however no Hth binding sites (CTGTCA, (Ryoo et al., 1999) within the IBS constructs.

The ideal experiment to prove that the observed repression is indeed mediated by Mirror would be to generate *mirror* loss-of-function clones in a Gbe-IBS-*lacZ* background and check whether β-galactosidase staining is recovered within the clone. Unfortunately all *mirror* and *Iroquois* alleles that were available were generated by excision of P-elements that left the *lacZ* gene behind. These alleles therefore still express β-galactosidase in an *Iroquois*-related pattern, making this sort of experiment hard to interpret. Therefore we decided to address the issue of specificity by ectopically expressing *mirror* in the ventral half of the eye and testing its effect on β-galactosidase expression.

5.4. Ectopic expression of *mirror* in the ventral half of the disc causes β-galactosidase repression

To ectopically express *mirror* in a tissue specific manner we made use of the GAL4: UAS system, which allows selective expression of the UAS construct in the pattern of the GAL4 line of choice. The expression pattern of the Gbe-IBS-*lacZ* line was restricted to the ventral half of the eye disc, mainly ahead of the morphogenetic furrow in the anterior-most part of the disc. Therefore for this experiment we had to select a GAL4 line that would drive expression in a pattern including this region.

Our selection of the Gal4 line was complicated by the fact that β-galactosidase expression in the Gbe-IBS-*lacZ* line could only be detected in the presence of two copies of the transgene. Since both our insertions are located on the second chromosome, we could not make use of a second chromosome driver line whilst maintaining a homozygous Gbe-IBS-*lacZ* status. We therefore decided to use *fringe-*

Gal4, a construct that drives expression in the ventral half of the eye disc and in the whole of the antenna (Figure 5.7A). The crossing scheme for this experiment is shown in Figure 5.7B. Flies of the IBS-*lacZ*; UAS-Mirr/*fng*-Gal4 genotype can be distinguished from their sibs (IBS-*lacZ*; UAS-Mirr/TM6B or IBS-*lacZ*;*fng*-Gal4/TM6B) by the lack of the *Tubby* (*Tb*) marker. This marker has a "tubby larvae" phenotype, allowing easy selection prior to dissection and staining. The *Tb* larvae were also stained for β-galactosidase as internal controls.

A caveat of this experiment that must be considered is that, as previously discussed, *mirror* is known to repress expression of *fringe*. This could in theory result in a situation whereby ectopic expression of *mirror* in the ventral half of the eye disc would downregulate the driver and hence counteract its own ectopic activation. If this were happening, there should not be any difference in the levels of β-galactosidase expression within the *fringe* domain. Our results show that this is not the case.

Overexpression of *mirror* is embryonic lethal (McNeill *et al.*, 1997). When the crosses were kept at 25° we did not obtain any third instar non-*Tb* larvae, indicating that embryos were indeed ectopically expressing *mirror*. To obtain progeny that would survive to the third instar larval stage crosses were kept at 18° C, a temperature at which the Gal4 protein is not active and then shifted to 25° to allow for Gal4 activity. I experimented with the timing of the temperature shift to define the best protocol for obtaining viable third instar larvae of the correct genotype. The best results were obtained when crosses were kept at 18° for 5-6 days and then shifted to 25° until wandering third instar larvae emerged. *Tb* larvae were the first ones to emerge and were by far more numerous than their non-*Tb* siblings, indicating that temporary ectopic expression of mirror is also affecting survival.

X-gal staining of eye-antennal discs dissected from *Tb* larvae (IBS-*lacZ*; UAS-Mirr/TM6B or IBS-*lacZ*;*fng*-Gal4/TM6B) showed that they had the same β-galactosidase pattern as the homozygous IBS-*lacZ* lines (Figure 5.7E). However the non-*Tb* siblings, which were ectopically expressing *mirror* in the *fringe* pattern, had lost expression of β-galactosidase in the ventral part of the disc (Figure 5.7C and D). This result indicated that Mirror can indeed act on the IBS *in vivo* and that this results in repression of the expression of the reporter gene.

It should be noted that ectopic expression of *mirror* using the *fringe*-Gal4 driver causes extensive malformation of the eye disc. It is possible that overexpression of Mirror in the ventral domain causes dorsalisation of the disc, which means that the effect shown above may be due to general patterning defects and loss of disc identity rather than just loss of β-galactosidase expression. For example generalised expression of *Iroquois* genes in the eye disc using an *eyeless*-Gal4 driver has been reported to cause a reduction in the size of the disc (Yang *et al.*, 1999). In this case however the size of the disc was not affected. The antennal disc on the other hand was severely reduced in size or missing altogether demonstrating that ectopic *mirror* expression in the antennal disc leads to severe defects in antennal development.

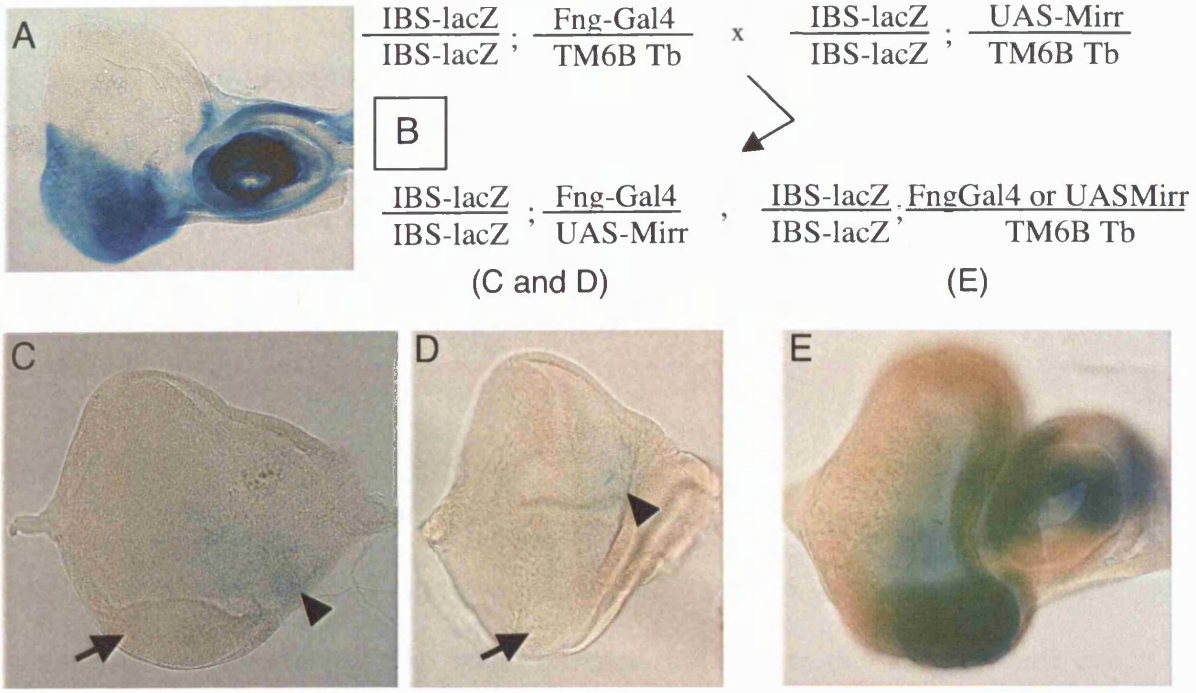


Figure 5.7: Ectopic Mirror expression in the ventral half of the eye represses Gbe-driven β-galactosidase. (A) X-gal staining of a *fringe*-lacZ line demonstrating the *fringe* expression pattern. (B) Outline of the cross scheme to drive ectopic expression of *mirror* in the ventral half of the eye disc using a *fringe*-Gal4 driver line. (C and D) Discs from larvae of IBS-lacZ; *fng*-Gal4/UAS-Mirr genotype (identified by the lack of the *Tb* marker). Expression of β-galactosidase is lost in the region where *mirror* is ectopically expressed (arrows). Note that some weak staining remains in other parts of the disc (arrowheads). (E) *Tb* siblings from the same cross exhibit the IBS-lacZ expression pattern (compare to figure 5.5)

5.5. Concluding remarks

The aim of this series of experiments was to test whether the IBS is functional *in vivo*. The effect of introducing four copies of the IBS in the background of a ubiquitous transcriptional activator was strong repression of β-galactosidase reporter expression in the domains of *Iroquois* expression both in the eye and in the wing imaginal disc.

We decided to focus our studies on the eye-antennal imaginal disc where the three fly *Iroquois* genes are expressed in the same pattern, allowing for an easier interpretation of the results. The extent of repression seen in the Gbe-IBS-*lacZ* eye discs suggests that there might be other proteins acting on the inserted element, though not necessarily on the IBS, thereby affecting transcription of the reporter gene, compared to the wing. The same is true for the Gbe-IBSmut-*lacZ* construct: the loss of dorsal repression could be interpreted as the effect of generating a site for a transcriptional activator expressed in the dorsal half of the eye disc. To our knowledge there are no binding sites for known transcriptional activators in the sequence of the mutant IBS. In any case, the fact that repression is extended to the ventral half of eye disc when Mirror is ectopically expressed in this domain strongly suggests that Mirror can act on the IBS to mediate transcriptional repression in the context of the eye imaginal disc.

I have also looked at the Gbe-IBS-*lacZ* expression in the wing disc where *Iroquois* genes are expressed in more divergent patterns. Similar to the effect seen in the eye, introducing the IBS tetramer in the background of the Grainyhead binding

element results in repression of β-galactosidase in the wing disc. Interestingly, mutating the IBS also results in loss of repression in the wing disc (data not shown), but due to the complexity of the expression patterns, further studies are required to assess the involvement of the *Iroquois* genes in this process.

Chapter 6:

An *in silico* approach for identification of *cis*-regulatory elements

6.1. Decoding the genome: genes and regulatory elements

Identifying regulatory elements within a genome is one of the most intriguing challenges in the post-genomic era. With an increasing number of whole genome sequences becoming available, the need to deduce the complete set of functional information, including genes, regulatory and structural elements is fundamental in handling and interpreting the bulk of emerging information. However this is far from achieved. Even for relatively small and compact genomes as that of the budding yeast (*Saccharomyces cerevisiae*) the number of true open reading frames (ORFs) is still debatable (Harrison *et al.*, 2002). The situation is even less clear for more complex genomes (Echols *et al.*, 2002).

Comparative analyses of the whole genome assemblies have shown that increased morphological and behavioural complexity cannot be accounted for by increase in gene numbers. The genome of the nematode worm *Caenorhabditis elegans* consists of nearly 20,000 genes (Ruvkun and Hobert, 1998) a number far greater than the approximately 13,000 that have been found in the *Drosophila* genome (Adams *et al.*, 2000). *Drosophila* nevertheless possesses a whole range of cell types and tissue structures that are not encountered in the nematode. Along the same lines, the

annotation of the human genome sequence has revealed that there are probably only around 30,000 genes in humans raising the question of what is the actual molecular basis of organismal complexity (Baltimore, 2001).

In principle there are different ways to exploit a relatively small-sized genome to achieve high levels of complexity. Alternative mRNA splicing (Graveley, 2001) and gene rearrangement (Agrawal *et al.*, 1998) have been shown to be successful strategies to increase the protein pool within a cell. Protein coding sequences, however, only represent a very small fraction of the typical metazoan genome, strikingly less than 2% in the case of humans (reviewed in Levine and Tjian, 2003). A far greater proportion of the genome is involved in control functions such as DNA replication, chromosome pairing and segregation and most importantly temporal and spatial regulation of gene expression.

In recent years there has been increasing evidence that regulation at the level of transcription is one of the major effectors of organismal diversity or, in other words, that morphological and behavioural complexity has relied on the evolution of differential regulatory mechanisms to control expression of a common set of genes.

6.1.1. What are enhancers?

The identification of regulatory elements is in many aspects far more challenging than that of genes. Enhancers are the most prevalent type of regulatory DNA sequences that determine when, where and at what level a given gene will be expressed during development. Enhancers were first described in relation to the beta-

globin genes (Banerji *et al.*, 1981). To transmit this level of positional and temporal information they need to integrate inputs from a variety of transcription factors and resolve them into an instructive output to the transcriptional machinery.

A typical enhancer element has an approximate length of 500bp and contains binding sites for two or more sequence specific transcription factors, activators and/or repressors. There have been different models as to what is the organization of an enhancer element. The prevalent idea is that enhancers serve as scaffolds that bring different combinations of transcription factors into close proximity and optimal arrangement to interact with the basal transcription machinery to switch target genes on or off (reviewed in Levine and Tjian, 2003).

Enhancers are found at various distances from the promoters of the genes they regulate, and currently there are examples of enhancers located in the 5', 3' or intronic sequences of targets genes. In some cases they are found to map closer to a promoter on which they have no effect, than to the promoter where their action is targeted on. In these cases there are various mechanisms to ensure that the enhancer interacts with the right promoter: DNA insulators (Burgess-Beusse *et al.*, 2002; Kellum and Schedl, 1991) are sequences first identified at gene boundaries that act to prevent *cis*-elements from one locus interfering with transcription in adjacent loci. Gene competition was first described in the chick globin locus and occurs when a shared enhancer preferentially interacts with one of the linked promoters (Choi and Engel, 1988). It is currently believed that this selectivity is dependent on the nature of *cis*-regulatory elements within the promoter so that some enhancers preferentially regulate TATA-containing promoters while others activate promoters containing other motifs such as Initiator sequences (INRs) or Downstream Promoter Elements (DPEs).

Despite the increasing amount of information that has been assembled on the structure and the special features of enhancer elements there is no obvious code allowing the prediction of regulatory elements based solely on sequence information. So far, most control elements have been identified by experimental manipulations, namely reporter assays with intact and mutated promoter and/or enhancer regions.

6.1.2. An *in silico* approach to identify Dorsal regulatory elements

One of the best systems currently used to study the organizational features of developmentally controlled enhancer elements is provided by the dorsal-ventral (DV) patterning of the *Drosophila* embryo. Dorsal is a sequence specific transcription factor that is expressed in a gradient along the DV axis of the early embryo with highest levels in the ventral regions and progressively lower distribution in lateral and dorsal regions (reviewed in Stathopoulos and Levine, 2002a).

Microarray screens have identified approximately 30 Dorsal targets that are responsive to different levels of the DV gradient (Stathopoulos and Levine, 2002b). The idea that enhancers that can respond to the same levels of Dorsal protein may share some common features initiated a detailed study aiming at deciphering a regulatory code linking primary DNA sequence with predicted patterns of gene expression (Markstein *et al.*, 2004).

It has previously been suggested that clustering of transcription factor binding sites is indicative of putative regulatory elements (Berman *et al.*, 2002). Searching the

genome for clusters of three or more Dorsal binding sites within a window of 400bp has identified 14 putative Dorsal elements that were subsequently tested in *in vivo* reporter assays. Only four of these were found to drive β-galactosidase expression along the DV axis in patterns similar to the associated endogenous genes. This shows that searching for clusters of sites for individual transcription factors is indicative but not very successful in identifying novel regulatory elements.

The success rate was significantly improved when known Dorsal elements were further dissected to identify binding sites for other transcription factors. These were then used to generate a template to search the genome for regions that shared these same features. In total only 7 clusters were identified that fulfilled the set criteria for nature, density and arrangement of binding sites. Three out of 7 corresponded to known Dorsal targets, two are likely to be false positives as they are associated with genes that have no Dorsal-related expression pattern and two are *bona fide* Dorsal targets.

These results show that computational methods can help to predict regulatory elements but the success rate remains relatively poor. The *in silico* analysis is still prone to false positives and has failed to predict some of Dorsal's true targets. The detailed characterization of more regulatory elements will increase the complement of transcription factor signatures that can be used to predict novel targets. The drawback of such an approach is that there is actually no proof that there should only be one "code" for all regulatory elements that responded to an individual or a group of transcription factors.

An additional tool to help discriminate between putative regulatory elements and false positives is currently becoming available with the completion of sequencing of more genomes. It is predicted that regulatory elements will be conserved amongst different species and scanning genomic regions from orthologous genes may provide additional information as to what are the important *cis*-elements for regulation of gene expression.

6.1.3. Can we use the IBS to identify Mirror targets *in silico?*

As discussed above in cases where there is evidence for direct targets of a transcription factor, a profile of the type, number and relative arrangement of transcription factor binding sites within enhancer elements can be used to identify new targets *in silico*.

In the case of Mirror or the rest of the Iroquois proteins in *Drosophila* there are no confirmed direct targets. One of the genes found to be downstream of Iroquois proteins is *fringe*. *fringe* has been shown in two different systems to be under the control of *mirror* although there is no evidence that this regulation is direct (Cho and Choi, 1998; Jordan *et al.*, 2000; Yang *et al.*, 1999). As discussed earlier clustering of transcription factor binding sites can be indicative of regulatory elements (Berman *et al.*, 2002; Markstein *et al.*, 2004). We arbitrarily chose to search within the 10kb upstream/downstream region of the *fringe* gene for clusters of IBSs. There are only two IBSs in this region, one at ~4kb upstream of the *fringe* transcription start and one downstream, but no IBS clusters neither 5'nor 3' of the *fringe* gene.

Clustering is nonetheless not a prerequisite for functional enhancers. Recent work by H. Jäckle's group has shown that the DNA fragments isolated from Krüppel associated chromatin by means of a Chromatin Immunoprecipitation assay are not enriched in clustered Krüppel binding sites (Matyash *et al.*, 2004). This means that even though the occurrence of clusters of binding sites may facilitate the *in silico* identification of true transcription factor targets, single occurrences of sites should by no means be disregarded.

I searched the *Drosophila* genome using Fly Enhancer (http://www.flyenhancer.com/) for occurrences of the ACAnnTGT motif and the results were as follows: there are ~43,000 occurrences of this site in the whole genome. Restricting the search to clusters of 3 sites within 400bp one comes up with ~5,000 occurrences, irrespective of their position in the genome. Using programmes like Fly Enhancer or Seqseek (http://flycompute.uoregon.edu/cgi-bin/seqseek.pl) one can obtain lists of *Drosophila* genes with specific number of sites located in selected upstream/downstream regions or within introns. These lists, although shorter, are still too extensive to test properly.

We are therefore combining this sort of *in silico* search with the results of a genome wide microarray approach undertaken by M. Mohns in the lab to identify direct Mirror targets. Many candidates for direct regulation can be selected based on both the microarray and the *in silico* analysis. Preliminary results have, however, shown that many of these genes are not true Mirror targets (Nadja Muncke, Mike Mohns and Helen McNeill, unpublished results). In the following sections an outline of the microarray and the case for a potential direct candidate will be discussed.

6.2. A microarray experiment to identify direct Mirror targets (M.Mohns)

For the purposes of the microarray experiment *mirror* was expressed under the control of a heat-shock promoter, and changes in gene expression levels were monitored after a 30 minute induction at 36° C using Affymetrix microarrays. To be able to subtract the effects of heat-shock on the general levels of gene expression control embryos were processed identically and examined in parallel. A diagram of the experimental settings is shown in Figure 6.1.

After 30 min of heat shock, expression of *mirror* increases in embryos carrying the *hs-mirr* transgene but not in control embryos subjected to the same heat-shock protocol. To avoid extremely high levels of overexpression, which may have non-specific effects various protocols have been tried as to the length and the temperature of the heat-shock, and the length of the recovery period after heat-shock. Under the above mentioned conditions (30 min, 36° C, no recovery period) only a moderate increase in *mirror* mRNA levels is induced indicating that embryos were probably exposed to *mirror* levels similar to those normally experienced *in vivo*.

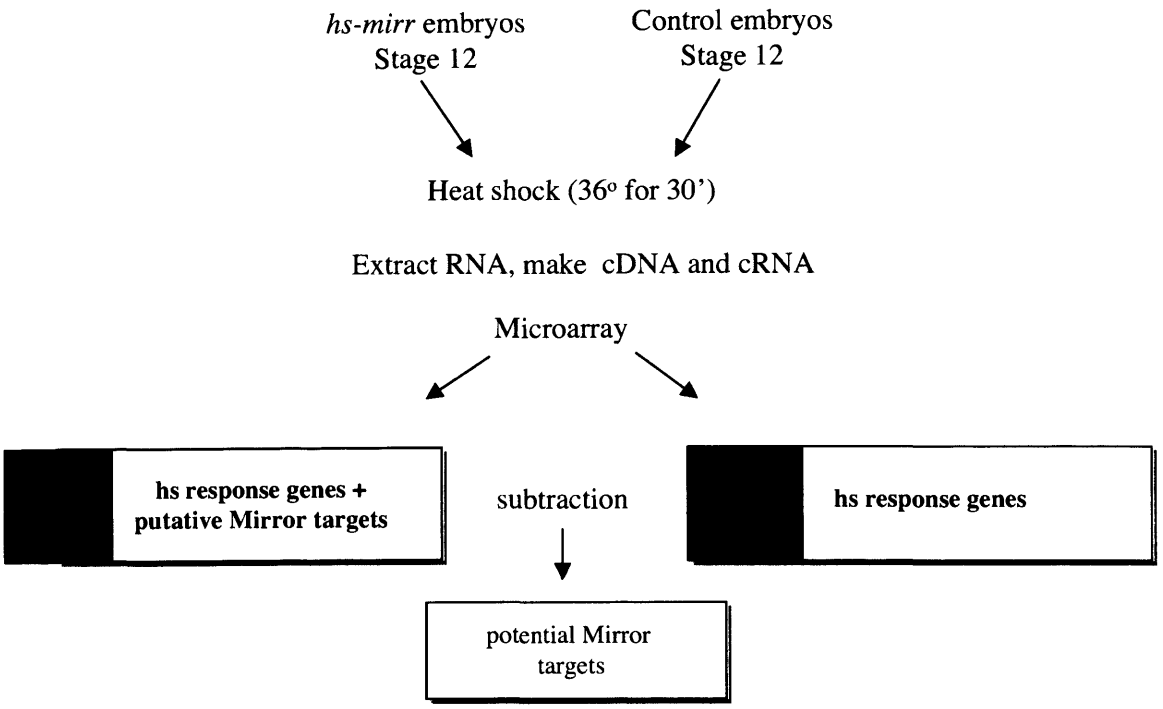


Figure 6.1: Outline of the microarray experiment (Mike Mohns). Total RNA was extracted from embryos expressing *mirror* under a heat-shock promoter or from control (yw) flies at the end of a 30 min heat-shock at 36°. Embryos were collected for 30 minutes and aged to stage 11 at room temperature before induction of the heat-shock. Independent samples were taken, processed in parallel and hybridized to Affymetrix microarrays. All experiments were carried out at least in triplicates. Synthesis of cDNA and labeled cRNA, array hybridization and scanning were performed according to Affymetrix protocols.

Amongst the genes whose levels were found to be affected by ectopic expression of mirror was *fringe*. *fringe* transcript levels were reduced rapidly within 30 min from the start of the heat shock. This reduction was not the most pronounced amongst the genes that showed a response to *mirror* overexpression. However together with the previously documented role of the *Iroquois* genes in regulation of *fringe* expression in various contexts (Cho and Choi, 1998; Jordan *et al.*, 2000; Yang *et al.*, 2002) this rapid response to *mirror* overexpression suggests that regulation may indeed be direct.

6.2.1. Identification of novel potential *mirror* targets using microarrays

Encouraged by the presence of *fringe* in the list of genes whose levels were rapidly affected by *mirror*, we went on to investigate other putative *mirror* targets. There was a substantial number of genes showing a specific response to Mirror overexpression and most of these were found to be downregulated. This was not surprising since previous reports on Iroquois members have shown that they act predominantly as transcriptional repressors (Cavodeassi *et al.*, 2001).

Unfortunately one of the greatest drawbacks of microarray analysis is that it is prone to a high rate of false positives (Freeman *et al.*, 2003). Several genes were fulfilling the statistical criteria for being considered as potential direct targets. The number however of genes confirmed to be affected by both Mirror overexpression and loss-of-function by *in situ* hybridisation was significantly lower (Mohns, 2003). I will

present below an example of a gene that, based on the microarray analysis, its experimentally documented response to *mirror* and the presence of IBSs in known or putative regulatory regions was selected for further investigation.

6.2.2. *Krüppel* is a candidate for direct *mirror* target

Krüppel is a zinc finger transcription factor mostly known for its role in embryonic segmentation, where it acts as a gap gene (Nusslein-Volhard and Wieschaus, 1980). At later stages *Krüppel* functions in malpighian tubule morphogenesis and formation of the embryonic CNS (Gaul and Weigel, 1990; Hoch *et al.*, 1990; Hoch *et al.*, 1991; Romani *et al.*, 1996). At stage 11 of embryonic development, when the microarray analysis was carried out, *mirror* is also involved in patterning of the CNS (Mohns *et al*, manuscript in preparation) and *mirror* and *Krüppel* have complementary expression patterns (Mohns, 2003) and Figure 6.2B-E). Heat shock induced over-expression of Mirror reduces the amount of *Krüppel* transcript as detected on Affymetrix microarrays (Figure 6.2A). Over-expression of UAS-*mirror* with the CNS specific or other drivers leads to severely reduced *Krüppel* expression in the embryonic CNS. This effect was not common to all neuroblasts suggesting that the role of *mirror* in *Krüppel* regulation is cell specific and that *mirror* is not the sole regulator of *Krüppel* expression in CNS development (Mohns, 2003).

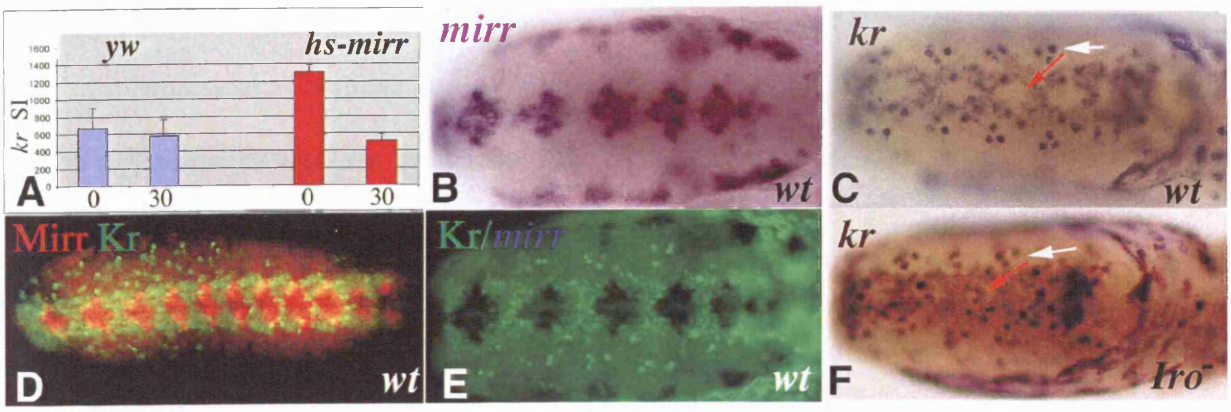


Figure 6.2: *krüppel* is a potential direct *mirror* target. Microarray data indicating that *krüppel* transcript is rapidly reduced in response to *mirror* overexpression after heat shock. Purple bars indicate Kr signal intensity (SI) in yw controls, red indicates expression in transgenic flies carrying a heat-shock driven *mirr* construct. (B) *mirr* transcript is expressed in a subset of cells in the embryonic CNS (stage 11). (C) Kr transcript is expressed in adjacent cells in the CNS at the same stage. (D) Double antibody staining for Kr and Mirr at stage 13 shows mostly non-overlapping expression. (E) Antibody and *in situ* staining for Kr and *mirr* respectively at stage 11 show non-overlapping expression. (F) *Iro* mutants ectopically express Kr in neuroblasts where *mirr* would normally be expressed (red arrows). Lateral cells are unaffected (white arrows). All pictures by Mike Mohns.

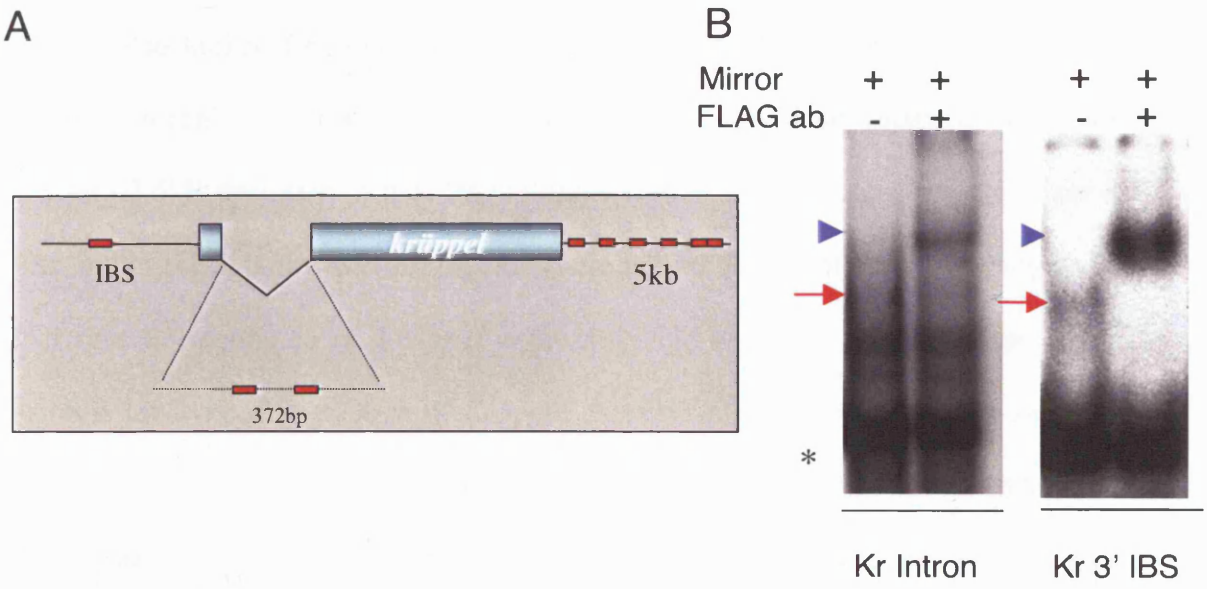


Figure 6.3: Mirror can bind IBS in the *krüppel (kr)* genomic region. (A) *kr* genomic structure indicating putative IBSs in red. Note the clustering of 6IBS within the 5kb downstream region. (B) Mirr can bind the IBSs located within the *kr* intron (left panel) and the 3' region (right panel) in EMSAs. Arrows indicate protein-DNA shifts and arrowheads the antibody supershifts. The asterisk indicates a non-specific band.

Consistent with this, loss of *Iroquois* function at the same stage resulted in ectopic *Krüppel* expression in at least two cells per segment, with the majority of embryos showing ectopic levels of *Krüppel* in most segments (Mohns, 2003) (Figure 6.3F). Since *mirror* is the only *Iroquois* gene expressed in the CNS we can attribute this effect to loss of *mirror*.

The *Krüppel* regulatory elements responsible for expression in the developing CNS have been extensively studied (Hoch *et al.*, 1990; Hoch *et al.*, 1991; Jacob *et al.*, 1991). We looked for clusters of IBSs and interestingly we can find two IBSs within the sole intron of the *Krüppel* gene (372bp) that has previously suggested to be involved in regulation of *Krüppel* expression in the CNS (Hoch *et al.*, 1990).

I also looked for occurrences of the IBS within the *Krüppel* upstream and downstream regions. There is 1 IBS within a large previously characterised CNS element ~7.6kB upstream of the *Krüppel* transcription start. Interestingly there are 6 IBSs within the 5kb downstream region. There are no data implicating this region in CNS specific regulation of *Krüppel* expression and we are planning to test if this region is involved in regulation of *Krüppel* expression by means of *in vivo* reporter assays. In order to test if Mirror can recognise these sites I performed EMSAs with labelled probes containing the IBSs from the intronic and the 3' *Krüppel* regions have shown that Mirror can bind these sites in the *in vitro* assay (Figure 6.3).

Besides *Krüppel,* other candidates from the microarray experiment were examined for the presence of IBS in putative or known regulatory elements and as expected several of these contained IBSs. Further analysis, including detailed study of

expression patterns, under wild-type, Mirror overexpression and loss-of function conditions as well as reporter analysis with wild-type and mutant genomic fragments is required to distinguish which of these genes correspond to true *mirror* targets.

6.3. Reporter analysis to validate potential mirror *targets.*

A classic assay for direct transcriptional control relies on the demonstration that the transcription factor of interest can act on a specific regulatory element from the candidate's genomic region to mediate transcriptional activation or repression. For this purpose I decided to generate various reporter constructs, in which I fused genomic fragments of the *fringe* and the *Krüppel* region to the *lacZ* reporter gene and generated transgenic flies.

In the case of *fringe* there are no data on *cis*-regulatory elements for any of the systems where *fringe* has been shown to repressed by *mirror*. I have cloned a 4.2kb genomic fragment lying upstream of the *fringe* transcription start and containing the 5' IBS in a *lacZ* reporter vector. This has been injected into embryos to generate transgenic flies. Because the *fringe* regulatory regions have not been studied before we did not make use of the Grainyhead vector. Our primary aim is to find out if this 4 kb fragment contains all necessary elements for spatially restricted *fringe* expression. If this is the case deletion analysis will be carried out to define what is the minimal element that can recapitulate the *fringe* expression pattern. Finally if this fragment contains the IBS, mutations will be introduced into the IBS and the effects will be assessed in reporter assays.

Preliminary experiments in the eye disc have shown that this fragment mediates a specific, but not strictly ventral, expression of the reporter gene. This implies that there may be more elements, located further upstream or downstream, that are additionally required to restrict expression of *fringe* in the ventral domain.

The same type of analysis will be carried out in the case of the *Krüppel* elements. I have already obtained transgenic lines that carry the 372p intron in a *lacZ* vector. Preliminary results obtained with one of these lines have shown β-galactosidase expression in a CNS–like expression pattern. Further evidence is required to define if the intronic IBSs are involved in *Krüppel* regulation. This will involve obtaining more lines with the same expression pattern as well as demonstrating that mutation of the IBSs within this construct alters the expression of β-galactosidase in a *mirror*-specific pattern. The 5kb downstream region will also be analysed in the same system.

6.4. Concluding remarks

In this chapter I present an outline of the current status of the *in silico* approaches for the identification of downstream targets based on the knowledge of binding specificities and in combination with some functional data (when these are available). Searching the whole genome for transcription factor binding sites generates, however, lists of sites and of their associated genes that are too extensive to test systematically.

We therefore chose to focus our studies on two potential downstream targets, one based on previous data on *mirror* function in several systems (*fringe*) and a novel candidate (*Krüppel*) identified through a microarray experiment recently performed in our lab (Mohns, 2003). If Mirror directly controls expression of these genes through the IBS then the presence of IBSs in the genomic regions around these genes should be indicative of *cis*-regulatory elements. There are no characterised *fringe* regulatory elements so for lack of other information our initial attempts will be solely guided by the presence of IBSs. For *Krüppel* on the other hand there is a complicated pattern of upstream and intronic sequences that are involved in regulation of its expression. Interestingly there are two IBSs within the single, short (372bp) intron of the *Krüppel* gene, previously suggested to be involved in CNS-specific expression (Hoch *et al.*, 1990). We have additionally found a 5kb uncharacterised region immediately downstream of the *Krüppel* ORF that contains 6IBSs. I have shown that Mirror binds the IBSs in the *Krüppel* intron and the 3' region in EMSAs (the most proximal part of the 3'UTR was used for the assay), confirming the idea that Mirror can recognise the IBS in the genomic context of putative regulatory elements.

We initiated an analysis of these putative regulatory elements by generating reporter constructs that will be tested for their ability to drive transcription in the pattern of the gene they are associated with and in response to Mirror. In combination with the ongoing analysis of other potential Mirror targets we hope to be able to demonstrate whether Mirror acts through the IBS to control expression of its downstream targets.

Chapter 7: Discussion

Iroquois proteins are a large family of transcription factors with homologues in a wide variety of evolutionary distant species, from sponges to humans. They are atypical Homeodomain transcription factors and share a well conserved Homeodomain of the TALE class and a novel domain of homology named the IRO-box with an as yet unidentified function. Outside these domains there is very little conservation. *Iroquois* were first studied in *Drosophila*, where the family consists of three genes: *mirror, ara* and *caup* (Gomez-Skarmeta *et al.,* 1996, McNeill *et al.*, 1997). *Drosophila Iroquois* have a very dynamic expression pattern and are involved in several processes from early embryonic stages till adulthood. In vertebrates there are 6 homologues, implying that a cluster duplication may have taken place to give rise to the full complement of *Iroquois* genes. Vertebrate *Iroquois* have been implicated in neural tube patterning, brain and heart development and axonal pathfinding in the retina (see also chapter 1.3.1).

In a rather simplified view their function can be summarised in that early in development they are expressed in broad domains and act to specify large territories while at later stages their expression becomes restricted to smaller regions within these domains affecting development of specific structures. This implies that their expression must be under a well-defined temporal and spatial control and that their downstream targets may vary based on the tissue and/or the stage of development. It has been reported that *ara* and *caup* may be redundant as their expression patterns are identical in all systems they have been described. *mirror* shares some of these

expression patterns but is in general more divergent both in structure and in expression.

The primary aim of my project was to understand Mirror's function at the level of transcription regulation. The first question I asked was what are Mirror's binding specificities (chapter 2 and 3). Unlike classic HOX proteins Mirror binds a novel motif consisting of a short palindromic sequence ACAnnTGT flanked by sequences rich in As and Ts. The presence of the AT-rich flanking regions is not essential for binding as Mirror can bind the ACAnnTGT palindrome in various contexts. However, the efficiency of binding, as measured by direct competition assays, is higher for the AT rich sequences.

Because of the palindromic nature of the motif, and the fact that the bHLH protein Myo D binds a very similar sequence (AACACGTGTT) forming homodimers we decided to test if Mirror would also bind DNA as a homodimer. EMSAs with a mixed population of Mirror proteins carrying different tags showed that the Mirror-DNA complexes could be super-supershifted when the two respective antibodies were included into the reaction, indicating the presence of at least two Mirror molecules in each complex. Pull down experiments have also suggested formation of a homodimer, although further analysis is required to identify the domains essential for formation of the complex.

Having established the binding specificities for Mirror I went on to test if these were conserved amongst other members of the family both in flies and in vertebrates. Based on the experimental results shown in chapter 4 *Drosophila* Ara and a vertebrate homologue (mouse Irx4) can bind the same site *in vitro*. It should be noted

that using equal amounts of labelled probe and equivalent amounts of protein (as verified by western blotting) binding of Mirror to the ACAnnTGT seemed to be more efficient than that of Ara and mIrx4 as evinced by the intensity of the bands in the EMSAs. *Iroquois* members share an almost identical HD and the DNA-binding third helix within the HD is perfectly conserved amongst all *Drosophila* members and also in the case of mouse Irx4. Outside the HD however the level of homology is very low (Ara and Caup are more similar to each other than they are to Mirr) suggesting that the overall conformation of the molecules may be variable which may in turn affect binding specificities. In all cases tested the affinity of Iroquois proteins was higher for the ACAnnTGT motif than for classic the HOX (ATTA) consensus as shown both by direct competition assays and by comparing the intensity of the bands when equivalent amounts of probe and protein were used.

Interestingly Mirror can also form heterodimers with Ara indicating that *Drosophila* Iroquois may in some cases act cooperatively to regulate downstream targets. Homo- or heterodimer formation is not a very common feature of HD transcription factors, there are however several cases where formation of dimeric or multimeric complexes has been reported (see also chapter 3.1.7). One of the best-studied cases is that of the PBC family. Like Iroquois proteins, members of the PBC family are atypical HD proteins and can bind DNA either on their own or in complexes with other HD proteins. Cooperative binding with PBC family members can shift the binding specificities of the complex towards a bipartite site (Chang *et al.*, 1996), or switch the transcriptional activity of the binding partner from activator to repressor (Pinsonneault *et al.*, 1997). Recently a new function has been assigned to a PBX/MEIS complex in the context of the activation of the myogenic pathway. Based on studies by Berkes and colleagues PBX/Meis complexes may act as "pioneer" proteins to

penetrate transcriptionally silent chromatin to mark specific genes for activation by the myogenic protein MyoD (Barkes *et al.*, 2004). According to this model, activation of the target genes depends on the recruitment of chromatin remodelling factors, which are as yet to be identified. Interestingly in a yeast–two-hybrid gene performed in our lab for Mirror binding partners one of the strongest interactors was the ATP dependent chromo-helicase CHD1 suggesting that Mirror may also act cooperatively with chromatin remodelling factors (Dahlsveen, 2002). Till now there is no reported interaction between Mirror (or any other Iroquois in flies or vertebrates) and members of the PBC and MEIS families. It has however been shown that quail Irx4 associates with the retinoic X receptor and that this interaction results in repression of the downstream target MyHC3 (Wang *et al.*, 2001) showing that Iroquois proteins are likely to act in complexes with other transcription factors.

The *in vitro* data indicated that Iroquois proteins bind a novel site and the best way to test if this site was functional, in other words capable of mediating transcriptional control *in vivo,* was to generate reporter constructs. Four repeats of the *in vitro* identified binding site were introduced into a lacZ vector that also carried sites for a transcriptional activator. This would allow for detection of transcription repression since, based on the literature, Iroquois have in many cases been reported to act to repress expression of target genes. In agreement to the previous reports Mirror binding to the [a/t]ACAnnTGT[a/t] site is capable of repressing reporter gene expression in the dorsal domain of the *Drosophila* eye disc where *Iroquois* genes are normally expressed. Mutation of the binding site alleviated the repression and ectopic expression of Mirror using the GAL4–UAS system extended the effect in domains where Iroquois are not normally expressed. These results indicated that the *in vitro* identified binding site was functional *in vivo.*

Many transcription factors bind *in vitro* to short, degenerate sequences that occur frequently in the genome. In the *Drosophila* genome there are around 43,000 occurrences of the minimal binding site ACAnnTGT, so identifying putative downstream targets solely based on the presence of binding sites is an impossible task. For this reason we decided to make use of other resources to support the *in vitro* data. One of the approaches that I initially took was an *in vivo* cross linking method to look for direct Mirror binding in fragments of chromatin treated with formaldehyde to stabilise all protein-DNA interactions. In this way binding of a protein to DNA can be studied in an *in vivo* context. Unfortunately the genomic regions that were then selected for analysis do not correspond to Mirror targets that later emerged from the microarray experiment, and therefore no useful conclusions could be made. The *in vivo* cross-linking method is however a very useful tool to study transcription factor binding in cells and could provide information as to how Mirror can discriminate *in vivo* between sites that *in vitro* would be bound equally well.

This would be an important step to identify and characterise downstream targets and provide a further insight into the role of Mirror in *Drosophila* development. The *in vitro* specificity of the newly identified binding site, the fact that it is recognised by other Iroquois proteins and, most importantly, its ability to act *in vivo* to control transcription, as shown by reporter assays in transgenic flies, strongly suggest that this site is part of the mechanism by which Mirror acts during development to control expression of target genes. In combination with the genome –wide microarray experiment aimed at identifying direct Mirror targets we have now a list of potential candidates that fulfil the appropriate requirements i.e. their levels are responsive to Mirror overexpression and loss-of-function and they have sites in putative regulatory

elements. We will be further testing these elements in reporter assays to identify the specific features that are necessary for Mirror function.

The identification of the first set of direct targets will mark a great advance in the study of the transcriptional activity of the Iroquois. It will allow us to establish an initial "signature" for Iroquois binding. This way one can take advantage of the power of the *in silico* analysis to identify sites for putative co-factors and therefore facilitate the identification and characterisation of novel regulatory elements. Co-occurrence of transcription factors binding sites has been suggested to be indicative of functional regulatory elements (Berman et al., 2002, Markstein et al., 2004) but this is not an indispensable requirement as single sites have also been shown to be transcriptionally functional (Matyash et al., 2004). Acquiring and consequently analysing the data on Iroquois binding within *in vivo* regulatory elements will provide valuable information on the function of this evolutionary conserved family of transcription factors.

Chapter 8: Materials and Methods

8.1.General DNA manipulation techniques

8.1.1.DNA preparation, restriction digestion and analysis

DNA preparation and purification was performed from bacterial cultures in Luria Bertani (LB) medium prepared by Cancer Research UK research services according to Sambrook *et al*, 1989. QIAGEN Mini and Maxi prep kits were used for purification according to the manufacturer's instructions. Restriction digestion was performed using restriction endonucleases from New England Biolabs (NEB) in the provided reaction buffers and under the conditions described in the NEB catalogue. Double digestions were performed as in the Double Digests section of the NEB catalogue.

For preparative and analytical gel electrophoresis of digested DNA 0.8-1.5% agarose gels in TAE were used (Sambrook *et al*, 1989, ultra-pure Electrophoresis grade agarose from Invitrogen).

For the purification of the digestion reactions from gels the QIAGEN Gel Extraction kit was used for fragments ranging from100bp-10kb while for fragments larger than 10kbs the QIAEX II kit was used to avoid shearing. Alternatively the digested fragments were separated from small fragments (<50bp) using the PCR

purification kit (QIAGEN). Ligation reactions were performed using T4 ligase (NEB) for 2-6 hours at room temperature or overnight at 16°C.

8.1.2. Transformation of competent bacterial strains

For transformation of ligated plasmids Invitrogen One Shot® Chemically Competent TOP10 cells {F⁻ *mcr*AΔ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *deo*R *ara*D139 *Δ(ara-leu)*7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G}were used according to manufacturers instructions. Alternatively I used TOP10 cells made competent using the Calcium Chloride method from Short Protocols in Molecular Biology (Ausubel, 1989). Transformed bacteria were plated on LB agar plates containing the appropriate antibiotics (SIGMA) at concentrations described in (Sambrook, 1989).

8.1.3. Sequencing

DNA Sequencing was carried out at the Cancer Research UK in house facilities using the ABS PRISM BigDye® Terminator Cycle Sequencing Kit and the Capillary Sequencing (PRISM 3730) kit. For each reaction 150-200ng of plasmid DNA, 3.2 pmol of the appropriate sequencing primer and 8µl of Terminator Ready Reaction Mix at a 1:16 dilution were used. The following Thermal cycling protocol was applied to all reactions:

Initial denaturation:

96°C	for 3 minutes

25 cycles as follows:

96°C	for 10 seconds
48°C	for 5 seconds
60°C	for 4 minutes

Reactions were stopped by adding 2µl of 125mM EDTA and were purified by Sodium Acetate/Ethanol precipitation. Sequencing results were analysed using the ABI PRISM and Sequencher softwares.

Sequencing primers used:

Primer	Sequence (5'-3')	Vectors
T7	TAATACGACTCACTATAGGG	pFTX9-FLAG pFTX11-HA
M13 (F)	GTAAAACGACGGCCAGT	pCR®-Blunt, pCR®2.1TOPO, pBluescript, PMT/V5-HIS TOPO
M13 (R)	CAGGAAACAGCTATGAC	pCR®-Blunt, pCR®2.1-TOPO, pBluescript, PMT/V5-HIS TOPO
HRE	GCGACGTGTTCACTTTGC	PBlueRabbit, pHZ50PL-Gbe
"Mirr 2" (I. Dahlsveen)	GTGGCCATGGCATGGCCCTGGCCATATCCA	Internal Mirr primer (upstream of HD)
"Seq4" (I. Dahlsveen)	ATGTCGGCCAGCGACCAGATCCTT	Internal Mirr primer (downstream of HD)

8.1.4. Polymerase Chain Reaction (PCR)

PCR reactions were carried out using Taq Polymerase (in house), Taq Master Mix (QIAGEN), Pfu Turbo Polymerase (Stratagene). For large, genomic fragments (>5kb) the ThermalAce DNA Polymerase Kit (Invitrogen) was used. Reactions were performed according to the manufacturers' instructions using plasmid- or genomic-DNA templates. Cycling was performed in a Primus 96plus PCR machine (MWG-BIOTECH) and a PTC-225 Peltier Thermal Cycler (MJC Research). PCR products generated with Taq polymerase reactions were used for direct cloning using the TOPO-TA Cloning Kit (Invitrogen). PCR products generated with Pfu and ThermalAce polymerases were incubated with Taq polymerase at 72 °C for 10minutes for post-amplification addition of 3'A-overhangs before proceeding to the TOPO-TA cloning step. PCR products were purified using the PCR purification Kit (QIAGEN).

PCR primers used

Primer	Sequence (5'-3')	Comments/Features
FLAG-F	GTCACTGTTGGGTCTTCTCTG	Subcloning of FLAG-Mirr from the pFTX9-FLAG into thePMTV5His vector for S2 cell expression
FLAG-R	CGGAAGCTTGCTTACATTTGC	Subcloning of FLAG-Mirr from the pFTX9-FLAG into thePMTV5His vector for S2 cell expression
AraF_NcoI	CCATGGCTGCCTACACACAATTC	For subcloning Ara from pBS to pFTX9FLAG. Introduces NcoI site at ORF start.
AraR	CTATATATATGGTACTATATCCG	For subcloning Ara from pBS to pFTX9FLAG.
5'NcoI	CACCCGCCATGGCCTACCCGC	For subcloning Irx4 from pBK CMV to pFTX9FLAG. Introduces NcoI at ORF start.
pBKCla3'	ACCCGGGTGGAAAATCGATGG	For subcloning Irx4 from pBK CMV to pFTX9FLAG. Introduces ClaI site downstream of cDNA.
F'MirHD-Nco	GGCCATGGATTTGAATGGGGCCAGAAGG	For subcloning Mirr HD into pFTX9FLAG. Introduces NcoI site 5' of HD.
R'MrrHD-Cla	GAATCGATAGGCTCCCAGGTCATCTTG	For subcloning Mirr HD into pFTX9FLAG. Introduces ClaI site 3' of HD.
Kr Intr L	CAATGCTTCAAGACGCACAAACG	For amplification of *Krüppel* intron from genomic DNA
Kr Intr R	GTTTTATGCCAGCTAATGCAG	For amplification of *Krüppel* intron from genomic DNA
Fng4kbFgreen with NotI	GCGGCCGCCCCTTCATATAGGGCAACAC TGACTC	For amplification of *Krüppel* intron from genomic DNA. Introduces NotI for subcloning in pBlueRabbit vector
Fng4kbR	ACGGTTACGGACCACTACGCGCACA	For amplification of *fringe* 4.2kb upstream region from genomic DNA
FLAG-F	CGGAAGCTTCGTTACATTTGC	For subcloning of FLAG-Mirr from pFTXP to pMT V5His (cell culture)
FLAG-R	GTCACTGTTGGGTCTTCTCTG	For subcloning of FLAG-Mirr from pFTXP to pMT V5His (cell culture)

8.1.5. Cloning

8.1.5.1 Cloning and expression vectors used:

Vector	Comments/ Source
pCR Blunt	PCR cloning vector , Invitrogen
pCR 2.1 TOPO-TA	PCR cloning vector, Invitrogen
pMTV5/His	PCR cloning vector for inducible expression in S2 cells, Invitrogen
pFTX9-FLAG	Expression vector for *in vitro* transcription/translation (T7 promoter) Gift from C. Hill (Howell et al., 1999)
pFTX11-HA	Expression vector for *in vitro* transcription/translation (T7 promoter) Gift from C. Hill (Howell et al., 1999).
pBlueRabbit	P-element insertion lacZ reporter vector for transgenics, gift from S. Bray
PHZ50PL-Gbe	P-element insertion lacZ reporter vector for transgenics, gift from S. Bray (Jennings *et al.*, 1999)

8.1.5.2. Subcloning of Mirror constructs into the pFTX9-FLAG vector

All Mirror constructs were previously subcloned into the pCR Blunt vector by PCR with NcoI and BglII sites inserted at beginning and end of ORF respectively by Ina Dahlsveen. The constructs shown in the table below were subcloned into the pFTX9-FLAG vector into NcoI and BamH1 sites. Bgl II and BamH1 produce compatible cohesive ends but the resulting product cannot be recleaved by either of the two enzymes. Diagnostic digestions were therefore performed with NcoI and XhoI (XhoI is within the MCS). Note that the pCR Blunt vector has 2 additional NcoI sites so that, prior to cloning, Mirror fragments were always purified based on their size on agarose gels.

Construct	Cut from pCR Blunt	Cloned into pFTX9	Size
Full length Mirror	Nco-BglII	Nco-BamH1	1937bp
Mirror Cterm	Nco-BglII	Nco-BamH1	1352bp
Mirror Cterm ΔHD	Nco-BglII	Nco-BamH1	923bp
Mirror Cterm ΔIRO	Nco-BglII	Nco-Bam H1	1910bp

8.1.5.3. Subcloning of Mirror cDNA into the pFTX11-HA vector

Full length Mirror cDNA was subcloned from the pFTX9 to the pFTX11 vector using NcoI-XhoI sites (NcoI site introduced by PCR at the beginning of the ORF, see above; XhoI site in the MCS).

8.1.5.4. Subcloning of *Drosophila* Ara and mouse Irx4 cDNA into the pFTX9 vector

Full length cDNA for *Drosophila* Ara in pBS (gift from J.L Gómez-Skarmeta) was PCR-amplified using the AraF_NcoI and AraR primers (see list of PCR primers) cloned into the pCR2.1 TOPO TA vector and then subcloned into the pFTX9-FLAG vector using the NcoI and XhoI sites.

Mouse Irx4 in pBK CMV (gift from V. Christoffels) was PCR amplified using the 5'NcoI and pBKCla3' primers (see list of PCR primers) cloned into the pCR2.1 TOPO TA vector and then subcloned into the pFTX9-FLAG vector using the NcoI and ClaI sites.

8.1.5.5. Cloning of a tetramer of the IBS and the IBSmut into the pHZ50PL-Gbe vector

Complementary single stranded oligonucleotides carrying 4 repeats of the IBS or the IBS mutant (for sequences see below) were synthesized at the Cancer Research UK oligonucleotide synthesis department. In each case oligos were annealed by mixing equal amounts of the two strands in 50 Mm KCl, heating at 100°C for 10 min and then slowly cooling down to room temperature. They were then digested with KpnI and NotI , ligated into the pHZ50 PL vector. DNA from transformed colonies was sequenced to test for possible oligomerisation and constructs that carried a single copy of the IBS (or IBS mutant) tetramer were injected into embryos (see below).

IBSKpnNot F' oligo:

5'GGCCGCAATTAACACGTGTTAATTGGTGGCTAATTAACACGTGTTAATTGGTGGCTAATT
AACACGTGTTAATTGGTGGCTAATTAACACGTGTTAATTGGTGGCTTGGTAC3'

IBSKpnNot R' oligo:

5'CAGCCACCAATTAACACGTGTTAATTAGCCACCAATTAACACGTGTTAATTAGCCACCAA
TTAACACGTGTTAATTAGCCAATTAACACGTGTTAATTGC3'

IBSmutKpnNot F' oligo:

5'GGCCGCAATTAAtACGTaTTAATTGGTGGCTAATTAAtACGTaTTAATTGGTGGCTAATTAAt
ACGTaTTAATTGGTGGCTAATTAAtACGTaTTAATTGGTGGCTTGGTAC3'

IBSmutKpnNot R' oligo:

5'CAGCCACCAATTAAtACGTaTTAATTAGCCACCAATTAAtACGTaTTAATTAGCCACCAATT
AAtACGTaTTAATTAGCCAATTAAtACGTaTTAATTGC3'

8.2. General protein manipulation techniques

8.2.1. In vitro transcription/translation reactions

In vitro synthesis of proteins was carried out using the TNT Quick Coupled Transcription/Translation rabbit reticulocyte lysate system (PROMEGA) according to the manufacturer's instructions.

8.2.2. *In vitro* transcription reaction

The reaction (100µl) was set as follows:

100mM rATP	1µl	1mM
100mM r CTP	1µl	1mM
100mM rUTP	1µl	1mM
10mM rGTP	1µl	0.1mM
5mM m7G(5')ppp(5')G	10µl	0.5mM
10X buffer 400mM Tris Cl pH 8.0 150mM $MgCl_2$	10µl	40mM Tris Cl pH 8.0, 15mM $MgCl_2$
1mg/ml template DNA	10µl	10µg
250mM DTT	5µl	12.5 mM
Rnasin (RNase inhibitor, PROMEGA) (60U/µl)	1µl	0.6U/µl
T7 polymerase	2µl	20 U/µl
Nuclease free H_20	58	

The reaction was incubated at 37°C for 30min. Then 1µl of 100mM rGTP was added and the reaction incubated for a further 60 minutes. RNA was purified using the QIAGEN RNeasy Kit and quantified spectrophotometrically. RNAs were then used for *in vitro* translation reactions

8.2.3. *In vitro* translation reaction

In vitro translation was carried out with the Rabbit Reticulocyte System , Nuclease Treated (PROMEGA) according to the manufacturer's instructions. For co-translations equal amount of RNAs were added into the reaction.

8.2.4. SDS-PAGE

Polyacrylamide gel electrophoresis was performed using the BioRad Protean III systems. Resolving gels were 7.5-10% polyacrylamide (37.5:1 acrylamide/Bis, BioRad) in 375mM Tris pH 8.8, 0.1% SDS while stacking gels were 4% polyacrylamide in 125mM Tris pH6.8, 0.1% SDS. Protein samples were mixed with 4X sample buffer (125mM Tris pH 6.8, 20% glycerol, 4% SDS, 10% β-mercaptoethanol, 0.008% Bromophenol Blue) and boiled for ~ 5min prior to loading. The BenchMark Molecular weight marker (Invitrogen) was used for approximate size determination. Gels were run in Tris-Glycine running buffer (25mM Tris pH8.3, 192mM Glycine, 0.1% SDS) at 200V for 30min-1hour.

8.2.5. Western blotting

Polyacrylamide gels and membranes (ECL-Hybond nitrocellulose; Amersham Pharmacia) were equilibrated in transfer buffer (48mM Tris pH8.3, 192 mM Glycine, 20% methanol, 0.05% SDS) for 10min. Transfer was performed using the Trans Blot Electrophoretic Transfer Cell (BIORAD) at 10V, overnight or the Semi Dry Electrophoretic Transfer Cell (BIORAD) at 10V for 1 hour. Membranes were blocked in PBS-Tween (PBS, 0.1% Tween-20) with 5% milk for 1 hour at room temperature

or overnight at 4°C. Primary antibodies were diluted in PBS-Tween, 2% milk and incubated with the membranes for 1-2 hours at room temperature. Membranes were washed 3x10min in PBS-Tween, 2%milk and then incubated with the appropriate HRP-conjugated secondary antibodies in the same buffer for 1 hour. Following 3 washes as before membranes were treated with the ECL Western Blotting chemiluminescent detection agent (Amersham Pharmacia) according to the manufactures instructions.

Primary antibodies used for Western blotting

Antibody	Source	Dilution	Comments
Rabbit α-C term Mirror	Helen McNeill	1:1000	Raised against GST fusion protein
rabbit α-Nterm Mirror (peptide)	Helen McNeill	1:500	Affinity purified against peptide
mouse α-FLAG (M2)	SIGMA	1:2000	Monoclonal
rabbit α-HA	Santa Cruz Biotechnology	1:1000	polyclonal

Secondary HRP-conjugated antibodies used for Western blotting

Antibody	Source	Dilution
α-mouse Ig HRP conjugate	Amersham Lifescience	1:2000
α-rabbit Ig HRP conjugate	Amersham Lifescience	1:5000

8.2.6 Immunoprecipitations

Immunoprecipitations with the α-FLAG M2 agarose beads were performed according to the manufacturer's instructions. For Western analysis no elution step was carried out and instead the beads were boiled in sample buffer (see section)

8.3. Electrophoretic Mobility Shift Assays (EMSAs)

8.3.1 Labelling of probes

Labelled nucleotides from Amersham (Redivue α-32P dCTP and α-32P dATP, 370MBq/ml -10mCi/ml). Probes were 25-30mers of double stranded DNA designed so that the F' and R' strands had a ~20bp overlap and carried 3'overhangs that could be filled in by Taq polymerase in the presence of α-^{32}P-dCTP. The labelling reaction was as follows:

F'oligo	100ng
R'oligo	100ng
PCR-buffer (No Mg^{2+}) 10X (Boehringer)	2μl
15mM $MgCl_2$	2μl
dATP, dTTP, dGTP (0.5mM each)	2μl
dCTP (40μM)	2μl
α-^{32}P-dCTP	3μl
Taq polymerase	1μl
ddH_2O	up to 20μl

Reactions were subjected to the following cycling scheme:

96°C for 1 min

15 cycles:

- 96°C for 30sec
- 50-55°C for 30sec (depending on the *Tm*, usually *Tm*-5°C)
- 72°C for 1min

72°C for 5min

Products were loaded on a 8% polyacrylamide (19:1 Acrylamide:Bis) gel in 0.5x TBE at 200V for 2h . After brief (~1min) exposure of an autoradiography film (KODAK, MXB Film) to allow for visualisation of the bands, probes were cut out and the DNA was eluted in 0.5mM NH_4Oac, 5mM EDTA and 0.1% SDS, overnight at 37°C. After phenol extraction and ethanol precipitation the probe was resuspended in dd H_2O and quantified by scintillation counting (Cerenkov counting). For all binding reactions 30,000 cpm of labelled probe were used.

8.3.2. Binding reactions and mobility shift gels

Binding reactions (30µl) were in: 25mM HEPES pH 7.5, 0.2mM EGTA, 0.2mM EDTA, 0.1M KCl, 0.5% NP-40, 1mM DTT, 600ng Poly(dIdC)-Poly(dIdC), 3mM $MgCl_2$, 5% Glycerol, 2µl reticulocyte lysate (*in vitro* translated protein) and 30,000cpm/reaction of labelled probe. The reaction was prepared in two steps: The probe mix (30,000cpm/reaction) was prepared in 600ng poly dI-dC, 3mM $MgCl_2$ (10µl reaction). The rest of the reagents including the *in vitro* translated protein (protein mix) were mixed to a final volume of 20µl. The two components (protein and probe mix) were then mixed. Complexes were allowed to form at room temperature for 20 min. For antibody supershifts 1µl of the α-FLAG, α-HA, α-Mirror antibodies were included in the protein mix prior to mixing with the probe. Reactions were run on a 4% polyacrylamide gels in 0.5x TBE at 200V for 2 hours. Before loading of the samples the gels were prerun under the same conditions for 2 hours.

8.3.3. DNA binding site selection assay

Selection of DNA sequences was carried out according to Pollock and Treisman, 1990 using the degenerate oligonucleotide R76 5'CAGGTCAGTTCAGCGGATCCTGTCG(N26)GAGGCGAATTCAGTGCAACTGCAGC3'. The oligo was labelled with [α^{32}P]-dCTP and [α^{32}P]-dATP by PCR using primer F: 5' CAGGTCAGTTCAGCGGATCCTGTCG3' and primer R: 5'GCTGCAGTTGCACTGAATTCGCCTC 3'. Mirror and FLAG-Mirror proteins were synthesized *in vitro* using the coupled transcription/translation rabbit reticulocyte lysate system (PROMEGA). Binding reactions (30µl) were in 25mMHEPES pH7.5, 0.2mM EDTA, 0.2mM EGTA, 100mM KCl, 0.1% NP40, 1mM DTT, 600ng poly (dIdC)-(dIdC), 3mM $MgCl_2$, 20% Glycerol. 2µl of the *in vitro* translation reaction were mixed with 0.2ng of labelled probe (for the first round 0.4ng of probe were used) in the presence of protease inhibitors (Roche). Protein–DNA complexes were allowed to form at room temperature for 20min and were immunoprecipitated with either an affinity purified rabbit α-Mirror peptide antibody bound to protein-A sepharose beads (AMERSHAM) or α-FLAG-coated agarose beads (SIGMA). DNA was recovered from the beads, amplified by PCR using primers F and R and used for a total of 4 rounds of selection. Selected oligos were subjected to EMSA. Bands appearing after 4 rounds of selection were excised, DNA was recovered, amplified and cloned using the TOPO-TA cloning system (Invitrogen). Inserts were sequenced and the sequences corresponding to the random core of the R76 oligo were analysed using the MEME (http://meme.sdsc.edu/meme/website/intro.html) and PWM (http://trantor.bioc.columbia.edu/Target_Explorer) softwares. Consensus sequences

were tested in mobility shifts assays with *in vitro* translated FLAG-Mirror and specific binding was verified by α-FLAG and α-Mirror supershifts.

8.3.4. List of oligonucleotides used for bandshifts

The sequence of the annealed oligonucleotides is given. This is the result of annealing two partially complementary single stranded oligos and filling in the overhangs by PCR.

Sequence of oligo	Comments/features
caggtttggagAAAAACACGTGTTAAgc	MEME and PWM consensus motif
ggtggctaACACGTGTtctgtgtgg	ACAnnTGT motif, GC spacer
ggtggctaACATATGTtctgtgtgg	ACAnnTGT motif, TA spacer
ggttctttgatcacttgACAtgTGTttgtgtgtgc	ACAnnTGT motif, TG spacer
ggtggctaAtACGTaTtctgtgtgg	Mutated motif
cgtttACATGTttttctccaaacctg	ACATGT (no spacer)
cgtttACAcTGTttttctcca	ACANTGT (1bp spacer)
cgtttACAcgcTGTttttctcca	ACA3NTGT (3bp spacer)
cgtttACAtataTGTttttctcccaaacctg	ACA4NTGT (4bp spacer)
cgtttACAtatataTGTttttctcccaaacctg	ACA6NTGT (6bp spacer)
	ACA8NTGT (8bp spacer)
ggtggctaACAtactgcatgaaTGTtctgtgtgg	ACA11NTGT (11bp spacer)
ggtggctaACAcgACActgtgtgg	ACA2NACA (direct repeats, 2N spacer)
ggtggctaACAcgaACActgtgtgg	ACA3NTGT (direct repeats, 3N spacer)
ggtggctacCAcgTGgtctattcggacctg	CANNTG (bHLH motif)
gctaagttaattaacacagaaatcaaattgc	L3 enhancer
gctaagttaattaaTacagaaatcaaattgc	L3 enhancer mutated in ACA
	L3 enhancer mutated in ATTA
ctagccattaatcagattaacggtgagcaattaga	DE goosecoid HOX motif
	P3 paired HOX motif
ggatagaaaatACAaaTGTaatgtaattgcacacataccg attagttagaatttgtttacatgtttggacaggaaccggcacttaa ctcgttatcgaccaaaacaaaaactagttagacgaaaatagag agctgcgaaaacactaagagttctctccgtacgaaactttctctc ACAcaTGTatcatatgt	*Krüppel* intron

8.4. Drosophila melanogaster techniques

8.4.1 Fly stocks

Stock name	Detailed genotype:	Source/comments
yw	*y ,w*	Ish-Horowicz lab
ry	*cn;ry*	Sarah Bray
Balancer (II)	*Sco*/SM6a, CyO	Ish-Horowicz lab
Balancer (III)	w; TM3, Sb, Ser/TM6 Tb	
Balancer (I)	Upd/FM7	Ish-Horowicz lab
Double balancer	w; *Sco/CyO;*MKRS Sb /TM3 Tb	Helen McNeill
fringe-Gal4	w; P{w+mW.hs=GawB} *fng*/TM3	Ken Irvine
mirr ^{e48}	*mirr^{e48} / TM3-lacZ*	Helen McNeill
UAS-*mirror* (pUAST-12)	*w^{1118}; UAS-mirr*	Helen McNeill homozygous viable (III)
Iro2	*Df(3L) IRO2/TM3-lacZ*	Juan Modolell
U4.33	P { ry^{+} Gbe-*lacZ*}	Sarah Bray (homozygous viable)
IBS 1-3	P { ry^{+} Gbe-IBS-*lacZ*} on the 2nd	homozygous viable
IBS 1-7	P { ry^{+} Gbe-IBS-*lacZ*} on the 2nd	homozygous viable
IBSmut 8a	P { ry^{+} Gbe-IBSmut-*lacZ*}	homozygous viable
IBSmut 19a	P { ry^{+} Gbe-IBSmut-*lacZ*}	homozygous viable
ry balancer (II)	Bc;Elp/CyO ; *ry*	Sarah Bray

Flies were grown in bottles and vials on a yeast-cornmeal-molasses-malt extract agar (sometimes including propionic acid or Nipagen) at 18°C, 25°C or room temperature (20-22°C).

8.4.2. P-element transformation

For the generation of the reporter gene transgenics the tetramer of the IBS, the IBSmutant tetramer and various fly genomic regions have been cloned into P-element transformation vectors (as described in the cloning section).

The P element plasmids pBlueRabbit and pHZ50PL-Gbe were co-precipitated either with the pTurbo helper plasmid in injection buffer (0.1mM Sodium Phosphate buffer, 5mMKCl) to a final concentration of 50 μg/ml. Alternatively a plasmid coding for the S129A activity mutant transposase (Beall *et al.*, 2002) was used (gift from K. Moses). For the co-precipitation a final concentration of 500μg/ml for both transforming plasmids was used.

DNA was injected into *yw* or *ry* embryos depending on the marker carried on the P-element plasmid (pBlueRabbit is a w^+ vector while pHZ50PL-Gbe is a ry^+ vector, gift from S. Bray). Injections were performed by Terence Gilbank and Steve Murray at Cancer Research UK and by Genetic Services, Boston, MA. Surviving adults were crossed back to *yw* or *ry* flies and transformant were selected by eye colour in the next generation. Single males were then used to set up crosses to balancer stocks (see list of fly-stocks) to identify the chromosome where the transgenes were inserted.

8.4.3. Xgal staining of eye-antennal discs

Third instar eye–antennal discs were dissected in PBS, fixed in 2% glutaraldehyde for 10 min at rt and incubated with 5-bromo-4-cloro-3-indolyl-β-galactopyranoside (X-Gal) at 37°C until detection of staining (usually o/n). Discs were mounted in 80%glycerol.

8.4.4 *Drosophila* tissue culture and transfection

Drosophila S2 cells (Invitrogen) were grown at 21-24°C in Schneider's Insect Medium (Sigma) supplemented with 10% foetal bovine serum (Invitrogen) and antibiotic/ antimycotic solution (SIGMA). Transfection was performed according to a Calcium-phosphate (CaPi) protocol (Pascal Meier). In general, 3 to 4.5 x 10^6 cells were seeded into each well of a 6 well plate and grown for 24 hours. A total of 5-7μg of DNA was added to a solution of 248mM $CaCl_2$ in a final volume of 125μl (Buffer A). This was then slowly added to 125μl of Buffer B (50mM HEPES, 1.5mM Na_2HPO_4, 280mM NaCl, pH 7.1) and was let to stand for 20min to allow for formation of the CaPi precipitate. Then, 200μl of the resulting precipitate was dropped slowly into each well of S2 cells. After 16 hours the transfection medium was replaced by fresh medium. Expression from the Metallothionein promoter was induced by the addition of $CuSO_4$ to a final concentration of 500μM 24 hours prior to harvest.

For the generation of stable lines the pMTV5His-FLAGMirr plasmid was transfected to S2 cells, according to the protocol from *Drosophila* Expression System

(Invitrogen). For the nuclear extracts a protocol adapted from Grant *et al*, 1992 was used. 6ml of confluent cells were spun at 5000rpm for 5min, then washed in ice-cold PBS and resuspended to HB buffer, transferred to eppendorfs and spun for 25sec at 4°C. They were then resuspended in lysis buffer and kept on ice for 10minutes. After a 5min spin at maximum speed (4°C) the pellet was resuspended in 200µl of Buffer C. It was then rocked vigorously for 15min on a rocking platform (4°C) and centrifuged one more time at 4°C for 5 minutes. The supernatant (nuclear extract) was frozen at-70°C.

HB Buffer: 10mM Tris, pH 7.3; 10mM KCl; 1.5mM $MgCl_2$; 0.5mM β-mercaptoethanol.

Lysis buffer: HB buffer + 0.4 % NP40

Buffer C: 20mM HEPES ph 7.9; 0.4mM NaCl; 1mM EDTA; 1mM DTT

Bibliography

Aboobaker, A. A., and Blaxter, M. L. (2003). Hox Gene Loss during Dynamic Evolution of the Nematode Cluster. Curr Biol *13*, 37-40.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-2195.

Affolter, M., Schier, A., and Gehring, W. J. (1990). Homeodomain proteins and the regulation of gene expression. Curr Opin Cell Biol *2*, 485-495.

Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature *394*, 744-751.

Akam, M. (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. Cell *57*, 347-349.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., *et al.* (1998). Zebrafish hox clusters and vertebrate genome evolution. Science *282*, 1711-1714.

Astell, C. R., Ahlstrom-Jonasson, L., Smith, M., Tatchell, K., Nasmyth, K. A., and Hall, B. D. (1981). The sequence of the DNAs coding for the mating-type loci of Saccharomyces cerevisiae. Cell *27*, 15-23.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1989). Short Protocols in Molecular Bioioly, John Wiley &Sons).

Axelrod, J. D., and McNeill, H. (2002). Coupling planar cell polarity signaling to morphogenesis. ScientificWorldJournal *2*, 434-454.

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Baltimore, D. (2001). Our genome unveiled. Nature *409*, 814-816.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299-308.

Bao, Z. Z., Bruneau, B. G., Seidman, J. G., Seidman, C. E., and Cepko, C. L. (1999). Regulation of chamber-specific gene expression in the developing heart by Irx4. Science *283*, 1161-1164.

Beall, E. L., Mahoney, M. B., and Rio, D. C. (2002). Identification and analysis of a hyperactive mutant form of Drosophila P-element transposase. Genetics *162*, 217-227.

Beeman, R. W. (1987). A homoeotic gene cluster in the red flour beetle. Nature *327*, 247-249.

Bellefroid, E. J., Kobbe, A., Gruss, P., Pieler, T., Gurdon, J. B., and Papalopulu, N. (1998). Xiro3 encodes a Xenopus homolog of the Drosophila Iroquois genes and functions in neural specification. Embo J *17*, 191-203.

Berkes C. A., Bergstrom D. A., Penn B. H., Seaver K. J., Knoepfler P. S. and Tapscott S.J. (2004) PBX marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol Cell *14*, 465-477.

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M., and Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. Proc Natl Acad Sci U S A *99*, 757-762.

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., and Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. Genes Dev *13*, 946-953.

Bertolino, E., Reimund, B., Wildt-Perinic, D., and Clerc, R. G. (1995). A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. J Biol Chem *270*, 31178-31188.

Billeter, M., and Wuthrich, K. (1993). Model studies relating nuclear magnetic resonance data with the three-dimensional structure of protein-DNA complexes. J Mol Biol *234*, 1094-1097.

Bosse, A., Stoykova, A., Nieselt-Struwe, K., Chowdhury, K., Copeland, N. G., Jenkins, N. A., and Gruss, P. (2000). Identification of a novel mouse Iroquois homeobox gene, Irx5, and chromosomal localisation of all members of the mouse Iroquois gene family. Dev Dyn *218*, 160-174.

Broadus, J., Skeath, J. B., Spana, E. P., Bossing, T., Technau, G., and Doe, C. Q. (1995). New neuroblast markers and the origin of the aCC/pCC neurons in the Drosophila central nervous system. Mech Dev *53*, 393-402.

Bruneau, B. G., Bao, Z. Z., Fatkin, D., Xavier-Neto, J., Georgakopoulos, D., Maguire, C. T., Berul, C. I., Kass, D. A., Kuroski-de Bold, M. L., de Bold, A. J., *et al.* (2001). Cardiomyopathy in Irx4-deficient mice is preceded by abnormal ventricular gene expression. Mol Cell Biol *21*, 1730-1736.

Bruneau, B. G., Bao, Z. Z., Tanaka, M., Schott, J. J., Izumo, S., Cepko, C. L., Seidman, J. G., and Seidman, C. E. (2000). Cardiac expression of the ventricle-specific homeobox gene Irx4 is modulated by Nkx2-5 and dHand. Dev Biol *217*, 266-277.

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. Proc Natl Acad Sci U S A *99 Suppl 4*, 16433-16437.

Burglin, T. (1994). Guidebook to the Homeobox Genes (Oxford, Oxford University Press).

Burglin, T. R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res *25*, 4173-4180.

Burglin, T. R. (1998). The PBC domain contains a MEINOX domain: coevolution of Hox and TALE homeobox genes? Dev Genes Evol *208*, 113-116.

Cagan, R. L., and Ready, D. F. (1989). Notch is required for successive cell decisions in the developing Drosophila retina. Genes Dev *3*, 1099-1112.

Calleja, M., Renaud, O., Usui, K., Pistillo, D., Morata, G., and Simpson, P. (2002). How to pattern an epithelium: lessons from achaete-scute regulation on the notum of Drosophila. Gene *292*, 1-12.

Cavodeassi, F., Diez Del Corral, R., Campuzano, S., and Dominguez, M. (1999). Compartments and organising boundaries in the Drosophila eye: the role of the homeodomain Iroquois proteins. Development *126*, 4933-4942.

Cavodeassi, F., Modolell, J., and Campuzano, S. (2000). The Iroquois homeobox genes function as dorsal selectors in the Drosophila head. Development *127*, 1921-1929.

Cavodeassi, F., Modolell, J., and Gomez-Skarmeta, J. L. (2001). The Iroquois family of genes: from body building to neural patterning. Development *128*, 2847-2855.

Cavodeassi, F., Rodriguez, I., and Modolell, J. (2002). Dpp signalling is a key effector of the wing-body wall subdivision of the Drosophila mesothorax. Development *129*, 3815-3823.

Chan, S. K., Popperl, H., Krumlauf, R., and Mann, R. S. (1996). An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. Embo J *15*, 2476-2487.

Chang, C. P., Brocchieri, L., Shen, W. F., Largman, C., and Cleary, M. L. (1996). Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. Mol Cell Biol *16*, 1734-1745.

Chen, R., Halder, G., Zhang, Z., and Mardon, G. (1999). Signaling by the TGF-beta homolog decapentaplegic functions reiteratively within the network of genes controlling retinal cell fate determination in Drosophila. Development *126*, 935-943.

Cho, K. O., and Choi, K. W. (1998). Fringe is essential for mirror symmetry and morphogenesis in the Drosophila eye. Nature *396*, 272-276.

Choi, K. W., and Benzer, S. (1994). Rotation of photoreceptor clusters in the developing Drosophila eye requires the nemo gene. Cell *78*, 125-136.

Choi, O. R., and Engel, J. D. (1988). Developmental regulation of beta-globin gene switching. Cell *55*, 17-26.

Christoffels, V. M., Keijser, A. G., Houweling, A. C., Clout, D. E., and Moorman, A. F. (2000). Patterning the embryonic heart: identification of five mouse Iroquois homeobox genes in the developing heart. Dev Biol *224*, 263-274.

Diez del Corral, R., Aroca, P., JL, G. m.-S., Cavodeassi, F., and Modolell, J. (1999). The Iroquois homeodomain proteins are required to specify body wall identity in Drosophila. Genes Dev *13*, 1754-1761.

Dahlsveen I.K. (2002), Identification and characterisation of proteins that interact with the *Drosophila* transcription factor Mirror, University of London.

Doe, C. Q. (2003). Hyper-Neuroblast Map.
http://www.neuro.uoregon.edu/doelab/nbintro.html

Dominguez, M., and de Celis, J. F. (1998). A dorsal/ventral boundary established by Notch controls growth and polarity in the Drosophila eye. Nature *396*, 276-278.
Duncan, I., and Montgomery, G. (2002a). E. B. Lewis and the bithorax complex: part I. Genetics *160*, 1265-1272.

Duncan, I., and Montgomery, G. (2002b). E. B. Lewis and the bithorax complex: part II. From cis-trans test to the genetic control of development. Genetics *161*, 1-10.

Echols, N., Harrison, P., Balasubramanian, S., Luscombe, N. M., Bertone, P., Zhang, Z., and Gerstein, M. (2002). Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. Nucleic Acids Res *30*, 2515-2523.

Fanto, M., Clayton, L., Meredith, J., Hardiman, K., Charroux, B., Kerridge, S., and McNeill, H. (2003). The tumor-suppressor and cell adhesion molecule Fat controls planar polarity via physical interactions with Atrophin, a transcriptional co-repressor. Development *130*, 763-774.

Fanto, M., and McNeill, H. (2004). Planar polarity from flies to vertebrates. J Cell Sci *117*, 527-533.

Fanto, M., Weber, U., Strutt, D. I., and Mlodzik, M. (2000). Nuclear signaling by Rac and Rho GTPases is required in the establishment of epithelial planar polarity in the Drosophila eye. Curr Biol *10*, 979-988.

Flegel, W. A., Singson, A. W., Margolis, J. S., Bang, A. G., Posakony, J. W., and Murre, C. (1993). Dpbx, a new homeobox gene closely related to the human proto-oncogene pbx1 molecular structure and developmental expression. Mech Dev *41*, 155-161.

Fortini, M. E., Lai, Z. C., and Rubin, G. M. (1991). The Drosophila zfh-1 and zfh-2 genes encode novel proteins containing both zinc-finger and homeodomain motifs. Mech Dev *34*, 113-122.

Freeman, M. (1997). Cell determination strategies in the Drosophila eye. Development *124*, 261-270.

Freeman, M. R., Delrow, J., Kim, J., Johnson, E., and Doe, C. Q. (2003). Unwrapping glial biology: Gcm target genes regulating glial development, diversification, and function. Neuron *38*, 567-580.

Furukubo-Tokunaga, K., Flister, S., and Gehring, W. J. (1993). Functional specificity of the Antennapedia homeodomain. Proc Natl Acad Sci U S A *90*, 6360-6364.

Garcia-Bellido, A. (1973). Developmental Compartmentalisation fo teh wing disc of *Drosophila*. Nat New Biol *245*, 251-253.

Gaul, U., and Weigel, D. (1990). Regulation of Kruppel expression in the anlage of the Malpighian tubules in the Drosophila embryo. Mech Dev *33*, 57-67.

Gehring, W. J., Affolter, M., and Burglin, T. (1994a). Homeodomain proteins. Annu Rev Biochem *63*, 487-526.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994b). Homeodomain-DNA recognition. Cell *78*, 211-223.

Germain, S., Howell, M., Esslemont, G. M., and Hill, C. S. (2000). Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. Genes Dev *14*, 435-451.

Gilbert S. (2000) Developmental Biology (6th Ed), Sinauer Associates, Inc.

Glavic, A., Gomez-Skarmeta, J. L., and Mayor, R. (2001). Xiro-1 controls mesoderm patterning by repressing Bmp-4 expression in the Spemann organizer. Dev Dyn *222*, 368-376.

Glavic, A., Maris Honore, S., Gloria Feijoo, C., Bastidas, F., Allende, M. L., and Mayor, R. (2004a). Role of BMP signaling and the homeoprotein Iroquois in the specification of the cranial placodal field. Dev Biol *272*, 89-103.

Glavic, A., Silva, F., Aybar, M. J., Bastidas, F., and Mayor, R. (2004b). Interplay between Notch signaling and the homeoprotein Xiro1 is required for neural crest induction in Xenopus embryos. Development *131*, 347-359.

Gomez-Skarmeta, J., de La Calle-Mustienes, E., and Modolell, J. (2001). The Wnt-activated Xiro1 gene encodes a repressor that is essential for neural development and downregulates Bmp4. Development *128*, 551-560.

Gomez-Skarmeta, J. L., Diez del Corral, R., de la Calle-Mustienes, E., Ferre-Marco, D., and Modolell, J. (1996). Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. Cell *85*, 95-105.

Gomez-Skarmeta, J. L., Glavic, A., de la Calle-Mustienes, E., Modolell, J., and Mayor, R. (1998). Xiro, a Xenopus homolog of the Drosophila Iroquois complex genes, controls development at the neural plate. Embo J *17*, 181-190.

Gomez-Skarmeta, J. L., and Modolell, J. (1996). araucan and caupolican provide a link between compartment subdivisions and patterning of sensory organs and veins in the Drosophila wing. Genes Dev *10*, 2935-2945.

Gomez-Skarmeta, J. L., and Modolell, J. (2002). Iroquois genes: genomic organization and function in vertebrate neural development. Curr Opin Genet Dev *12*, 403-408.

Gomez-Skarmeta, J. L., Rodriguez, I., Martinez, C., Culi, J., Ferres-Marco, D., Beamonte, D., and Modolell, J. (1995). Cis-regulation of achaete and scute: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. Genes Dev *9*, 1869-1882.

Goriely, A., Diez del Corral, R., and Storey, K. G. (1999). c-Irx2 expression reveals an early subdivision of the neural plate in the chick embryo. Mech Dev *87*, 203-206.

Goutte, C., and Johnson, A. D. (1993). Yeast a1 and alpha 2 homeodomain proteins form a DNA-binding activity with properties distinct from those of either protein. J Mol Biol *233*, 359-371.

Graveley, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. Trends Genet *17*, 100-107.

Grillenzoni, N., van Helden, J., Dambly-Chaudiere, C., and Ghysen, A. (1998). The iroquois complex controls the somatotopy of Drosophila notum mechanosensory projections. Development *125*, 3563-3569.

Hanes, S. D., and Brent, R. (1989). DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. Cell *57*, 1275-1283.

Harrison, P. M., Kumar, A., Lang, N., Snyder, M., and Gerstein, M. (2002). A question of size: the eukaryotic proteome and the problems in defining it. Nucleic Acids Res *30*, 1083-1090.

Heberlein, U., Wolff, T., and Rubin, G. M. (1993). The TGF beta homolog dpp and the segment polarity gene hedgehog are required for propagation of a morphogenetic wave in the Drosophila retina. Cell *75*, 913-926.

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G., and et al. (1988). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. Genes Dev *2*, 1513-1516.

Hirsch, J. A., and Aggarwal, A. K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. Embo J *14*, 6280-6291.

Hoch, M., Schroder, C., Seifert, E., and Jackle, H. (1990). cis-acting control elements for Kruppel expression in the Drosophila embryo. Embo J *9*, 2587-2595.

Hoch, M., Seifert, E., and Jackle, H. (1991). Gene expression mediated by cis-acting sequences of the Kruppel gene in response to the Drosophila morphogens bicoid and hunchback. Embo J *10*, 2267-2278.

Houweling, A. C., Dildrop, R., Peters, T., Mummenhoff, J., Moorman, A. F., Ruther, U., and Christoffels, V. M. (2001). Gene and cluster-specific expression of the Iroquois family members during mouse development. Mech Dev *107*, 169-174.

Howell, M., Itoh, F., Pierreux, C. E., Valgeirsdottir, S., Itoh, S., ten Dijke, P., and Hill, C. S. (1999). Xenopus Smad4beta is the co-Smad component of developmentally regulated transcription factor complexes responsible for induction of early mesodermal genes. Dev Biol *214*, 354-369.

Irvine, K. D. (1999). Fringe, Notch, and making developmental boundaries. Curr Opin Genet Dev *9*, 434-441.

Itoh, M., Kudoh, T., Dedekian, M., Kim, C. H., and Chitnis, A. B. (2002). A role for iro1 and iro7 in the establishment of an anteroposterior compartment of the ectoderm adjacent to the midbrain-hindbrain boundary. Development *129*, 2317-2327.

Jacob, Y., Sather, S., Martin, J. R., and Ollo, R. (1991). Analysis of Kruppel control elements reveals that localized expression results from the interaction of multiple subelements. Proc Natl Acad Sci U S A *88*, 5912-5916.

Jarman, A. P., Grell, E. H., Ackerman, L., Jan, L. Y., and Jan, Y. N. (1994). Atonal is the proneural gene for Drosophila photoreceptors. Nature *369*, 398-400.

Jarman, A. P., Sun, Y., Jan, L. Y., and Jan, Y. N. (1995). Role of the proneural gene, atonal, in formation of Drosophila chordotonal organs and photoreceptors. Development *121*, 2019-2030.

Jennings, B. H., Tyler, D. M., and Bray, S. J. (1999). Target specificities of Drosophila enhancer of split basic helix-loop-helix proteins. Mol Cell Biol *19*, 4600-4610.

Jin, Z., Zhang, J., Klar, A., Chedotal, A., Rao, Y., Cepko, C. L., and Bao, Z. Z. (2003). Irx4-mediated regulation of Slit1 expression contributes to the definition of early axonal paths inside the retina. Development *130*, 1037-1048.

Johnson, F. B., Parker, E., and Krasnow, M. A. (1995). Extradenticle protein is a selective cofactor for the Drosophila homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. Proc Natl Acad Sci U S A *92*, 739-743.

Jordan, K. C., Clegg, N. J., Blasi, J. A., Morimoto, A. M., Sen, J., Stein, D., McNeill, H., Deng, W. M., Tworoger, M., and Ruohola-Baker, H. (2000). The homeobox gene mirror links EGF signalling to embryonic dorso-ventral axis formation through notch activation. Nat Genet *24*, 429-433.

Kasahara, H., Usheva, A., Ueyama, T., Aoki, H., Horikoshi, N., and Izumo, S. (2001). Characterization of homo- and heterodimerization of cardiac Csx/Nkx2.5 homeoprotein. J Biol Chem *276*, 4570-4580.

Kehl, B. T., Cho, K. O., and Choi, K. W. (1998). mirror, a Drosophila homeobox gene in the Iroquois complex, is required for sensory organ and alula formation. Development *125*, 1217-1227.

Keleman, K., Rajagopalan, S., Cleppien, D., Teis, D., Paiha, K., Huber, L. A., Technau, G. M., and Dickson, B. J. (2002). Comm sorts robo to control axon guidance at the Drosophila midline. Cell *110*, 415-427.

Kellum, R., and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. Cell *64*, 941-950.

Klein, T. (2001). Wing disc development in the fly: the early stages. Curr Opin Genet Dev *11*, 470-475.

Kobayashi, D., Kobayashi, M., Matsumoto, K., Ogura, T., Nakafuku, M., and Shimamura, K. (2002). Early subdivisions in the neural plate define distinct competence for inductive signals. Development *129*, 83-93.

Kophengnavong, T., Michnowicz, J. E., and Blackwell, T. K. (2000). Establishment of distinct MyoD, E2A, and twist DNA binding specificities by different basic region-DNA conformations. Mol Cell Biol *20*, 261-272.

Kudoh, T., and Dawid, I. B. (2001). Role of the iroquois3 homeobox gene in organizer formation. Proc Natl Acad Sci U S A *98*, 7852-7857.

Kumar, J. P. (2001). Signalling pathways in Drosophila and vertebrate retinal development. Nat Rev Genet *2*, 846-857.

Kumar, J. P., and Moses, K. (2001a). EGF receptor and Notch signaling act upstream of Eyeless/Pax6 to control eye specification. Cell *104*, 687-697.

Kumar, J. P., and Moses, K. (2001b). The EGF receptor and notch signaling pathways control the initiation of the morphogenetic furrow during Drosophila eye development. Development *128*, 2689-2697.

Kurata, S., Go, M. J., Artavanis-Tsakonas, S., and Gehring, W. J. (2000). Notch signaling and the determination of appendage identity. Proc Natl Acad Sci U S A *97*, 2117-2122.

Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell *66*, 305-315.

Lee, H. H., and Frasch, M. (2000). Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. Development *127*, 5497-5508.

Leiting, B., De Francesco, R., Tomei, L., Cortese, R., Otting, G., and Wuthrich, K. (1993). The three-dimensional NMR-solution structure of the polypeptide fragment 195-286 of the LFB1/HNF1 transcription factor from rat liver comprises a nonclassical homeodomain. Embo J *12*, 1797-1803.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565-570.

Leyns, L., Gomez-Skarmeta, J. L., and Dambly-Chaudiere, C. (1996). iroquois: a prepattern gene that controls the formation of bristles on the thorax of Drosophila. Mech Dev *59*, 63-72.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. Science *270*, 262-269.

Lim, J., and Choi, K. W. (2003). Bar homeodomain proteins are anti-proneural in the Drosophila eye: transcriptional repression of atonal by Bar prevents ectopic retinal neurogenesis. Development *130*, 5965-5974.

Lu, Q., and Kamps, M. P. (1996). Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: proposal for a model of a Pbx1-Hox-DNA complex. Mol Cell Biol *16*, 1632-1640.

Ma, P. C., Rould, M. A., Weintraub, H., and Pabo, C. O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell *77*, 451-459.

Mann, R. S., and Chan, S. K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet *12*, 258-262.

Markstein, M., Zinzen, R., Markstein, P., Yee, K. P., Erives, A., Stathopoulos, A., and Levine, M. (2004). A regulatory code for neurogenic gene expression in the Drosophila embryo. Development *131*, 2387-2394.

Matsumoto, K., Nishihara, S., Kamimura, M., Shiraishi, T., Otoguro, T., Uehara, M., Maeda, Y., Ogura, K., Lumsden, A., and Ogura, T. (2004). The prepattern transcription factor Irx2, a target of the FGF8/MAP kinase cascade, is involved in cerebellum formation. Nat Neurosci *7*, 605-612.

Matyash, A., Chung, H. R., and Jackle, H. (2004). Genome-wide mapping of in vivo targets of the Drosophila transcription factor Kruppel. J Biol Chem *279*, 30689-30696.

Maurel-Zaffran, C., and Treisman, J. E. (2000). pannier acts upstream of wingless to direct dorsal eye disc development in Drosophila. Development *127*, 1007-1016.

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., and Gehring, W. J. (1984a). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell *37*, 403-408.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283-302.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., and Gehring, W. J. (1984b). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature *308*, 428-433.

McNeill, H., Yang, C. H., Brodsky, M., Ungos, J., and Simon, M. A. (1997). mirror encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the Drosophila eye. Genes Dev *11*, 1073-1082.

Mead, P. E., Brivanlou, I. H., Kelley, C. M., and Zon, L. I. (1996). BMP-4-responsive regulation of dorsal-ventral patterning by the homeobox protein Mix.1. Nature *382*, 357-360.

Mendel, D. B., Hansen, L. P., Graves, M. K., Conley, P. B., and Crabtree, G. R. (1991). HNF-1 alpha and HNF-1 beta (vHNF-1) share dimerization and homeo domains, but not activation domains, and form heterodimers in vitro. Genes Dev *5*, 1042-1056.

Mohns, M. (2003) Identification of downstream targets of Mirror, a transcription factor in *Drosophila melanogaster*, University of London.

Morisato, D., and Anderson, K. V. (1995). Signaling pathways that establish the dorsal-ventral pattern of the Drosophila embryo. Annu Rev Genet *29*, 371-399.

Netter, S., Fauvarque, M. O., Diez del Corral, R., Dura, J. M., and Coen, D. (1998). white+ transgene insertions presenting a dorsal/ventral pattern define a single cluster of homeobox genes that is silenced by the polycomb-group proteins in Drosophila melanogaster. Genetics *149*, 257-275.

Nourse, J., Mellentin, J. D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S. D., and Cleary, M. L. (1990). Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell *60*, 535-545.

Nover, L., and Scharf, K. D. (1997). Heat stress proteins and transcription factors. Cell Mol Life Sci *53*, 80-103.

Nusslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795-801.

Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J., and Wuthrich, K. (1990). Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution. Embo J *9*, 3085-3092.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature *397*, 714-719.

Peifer, M., and Wieschaus, E. (1990). Mutations in the Drosophila gene extradenticle affect the way specific homeo domain proteins regulate segmental identity. Genes Dev *4*, 1209-1223.

Peters, T., Dildrop, R., Ausmeier, K., and Ruther, U. (2000). Organization of mouse Iroquois homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development. Genome Res *10*, 1453-1462.

Pichaud, F., and Casares, F. (2000). homothorax and iroquois-C genes are required for the establishment of territories within the developing eye disc. Mech Dev *96*, 15-25.

Pinsonneault, J., Florence, B., Vaessin, H., and McGinnis, W. (1997). A model for extradenticle function as a switch that changes HOX proteins from repressors to activators. Embo J *16*, 2032-2042.

Poellinger, L., and Roeder, R. G. (1989). Octamer transcription factors 1 and 2 each bind to two different functional elements in the immunoglobulin heavy-chain promoter. Mol Cell Biol *9*, 747-756.

Pollock, R., and Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res *18*, 6197-6204.

Postlethwait, J. H., and Schneiderman, H. A. (1971). A clonal analysis of development in Drosophila melanogaster: morphogenesis, determination, and growth in the wild-type antenna. Dev Biol *24*, 477-519.

Prince, V. E., and Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet *3*, 827-837.

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wuthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell *59*, 573-580.

Qian, Y. Q., Otting, G., Furukubo-Tokunaga, K., Affolter, M., Gehring, W. J., and Wuthrich, K. (1992). NMR structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the Drosophila Antennapedia protein. Proc Natl Acad Sci U S A *89*, 10738-10742.

Reifegerste, R., and Moses, K. (1999). Genetics of epithelial polarity and pattern in the Drosophila retina. Bioessays *21*, 275-285.

Reiser, L., Modrusan, Z., Margossian, L., Samach, A., Ohad, N., Haughn, G. W., and Fischer, R. L. (1995). The BELL1 gene encodes a homeodomain protein involved in pattern formation in the Arabidopsis ovule primordium. Cell *83*, 735-742.

Riechmann, V., and Ephrussi, A. (2001). Axis formation during Drosophila oogenesis. Curr Opin Genet Dev *11*, 374-383.

Rieckhof, G. E., Casares, F., Ryoo, H. D., Abu-Shaar, M., and Mann, R. S. (1997). Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. Cell *91*, 171-183.

Romani, S., Jimenez, F., Hoch, M., Patel, N. H., Taubert, H., and Jackle, H. (1996). Kruppel, a Drosophila segmentation gene, participates in the specification of neurons and glial cells. Mech Dev *60*, 95-107.

Ruvkun, G., and Hobert, O. (1998). The taxonomy of developmental control in Caenorhabditis elegans. Science *282*, 2033-2041.

Ryoo, H. D., Marty, T., Casares, F., Affolter, M., and Mann, R. S. (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. Development *126*, 5137-5148.

Sambrook, J., Fritsch, E. F. and Maniatis, T (1989). Molecular Cloning: a Laboratory Manual., Cold Spring Harbor Laboratory Press).

Schier, A. F., and Gehring, W. J. (1992). Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. Nature *356*, 804-807.

Seeger, M., Tear, G., Ferres-Marco, D., and Goodman, C. S. (1993). Mutations affecting growth cone guidance in Drosophila: genes necessary for guidance toward or away from the midline. Neuron *10*, 409-426.

Skeath, J. B. (1999). At the nexus between pattern formation and cell-type specification: the generation of individual neuroblast fates in the Drosophila embryonic central nervous system. Bioessays *21*, 922-931.

Sosinsky, A., Bonin, C. P., Mann, R. S., and Honig, B. (2003). Target Explorer: An automated tool for the identification of new target genes for a specified set of transcription factors. Nucleic Acids Res *31*, 3589-3592.

Stathopoulos, A., and Levine, M. (2002a). Dorsal gradient networks in the Drosophila embryo. Dev Biol *246*, 57-67.

Stathopoulos, A., and Levine, M. (2002b). Whole-genome expression profiles identify gene batteries in Drosophila. Dev Cell *3*, 464-465.

Steelman, S., Moskow, J. J., Muzynski, K., North, C., Druck, T., Montgomery, J. C., Huebner, K., Daar, I. O., and Buchberg, A. M. (1997). Identification of a conserved family of Meis1-related homeobox genes. Genome Res *7*, 142-156.

Strutt, D. I., Weber, U., and Mlodzik, M. (1997). The role of RhoA in tissue polarity and Frizzled signalling. Nature *387*, 292-295.

Tomlinson, A. (2003). Patterning the peripheral retina of the fly: decoding a gradient. Dev Cell *5*, 799-809.

Ton, C. C., Hirvonen, H., Miwa, H., Weil, M. M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N. D., Meijers-Heijboer, H., Drechsler, M., and et al. (1991). Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. Cell *67*, 1059-1074.

Treisman, J. E., and Heberlein, U. (1998). Eye development in Drosophila: formation of the eye field and control of differentiation. Curr Top Dev Biol *39*, 119-158.

Vershon, A. K., and Johnson, A. D. (1993). A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. Cell *72*, 105-112.

Vollbrecht, E., Veit, B., Sinha, N., and Hake, S. (1991). The developmental gene Knotted-1 is a member of a maize homeobox gene family. Nature *350*, 241-243.

Wang, G. F., Nikovits, W., Jr., Bao, Z. Z., and Stockdale, F. E. (2001). Irx4 forms an inhibitory complex with the vitamin D and retinoic X receptors to regulate cardiac chamber-specific slow MyHC3 expression. J Biol Chem *276*, 28835-28841.

Weber, U., Paricio, N., and Mlodzik, M. (2000). Jun mediates Frizzled-induced R3/R4 cell fate distinction and planar polarity determination in the Drosophila eye. Development *127*, 3619-3629.

Wernet, M. F., Labhart, T., Baumann, F., Mazzoni, E. O., Pichaud, F., and Desplan, C. (2003). Homothorax switches function of Drosophila photoreceptors from color to polarized light sensors. Cell *115*, 267-279.

Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., and Desplan, C. (1993). Cooperative dimerization of paired class homeo domains on DNA. Genes Dev *7*, 2120-2134.

Wolff, T., and Ready, D. F. (1993). Pattern formation in the Drosophila retina. In The development of *Drsosphila melanogaster*, M. B. a. A. Martinez-Arias, ed. (Clod Spring Harbor Press), pp. 1277-1325.

Yang, C. H., Axelrod, J. D., and Simon, M. A. (2002). Regulation of Frizzled by fat-like cadherins during planar polarity signaling in the Drosophila compound eye. Cell *108*, 675-688.

Yang, C. H., Simon, M. A., and McNeill, H. (1999). mirror controls planar polarity and equator formation through repression of fringe expression and through control of cell affinities. Development *126*, 5857-5866.

Zecca, M., and Struhl, G. (2002a). Control of growth and patterning of the Drosophila wing imaginal disc by EGFR-mediated signaling. Development *129*, 1369-1376.

Zecca, M., and Struhl, G. (2002b). Subdivision of the Drosophila wing imaginal disc by EGFR-mediated signaling. Development *129*, 1357-1368.

Zeidler, M. P., Perrimon, N., and Strutt, D. I. (1999a). The four-jointed gene is required in the Drosophila eye for ommatidial polarity specification. Curr Biol *9*, 1363-1372.

Zeidler, M. P., Perrimon, N., and Strutt, D. I. (1999b). Polarity determination in the Drosophila eye: a novel role for unpaired and JAK/STAT signaling. Genes Dev *13*, 1342-1353.

Zhao, D., Woolner, S., and Bownes, M. (2000). The Mirror transcription factor links signalling pathways in Drosophila oogenesis. Dev Genes Evol *210*, 449-457.